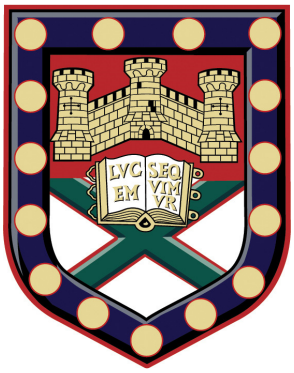
An investigation into dynamic and functional properties of prokaryotic signalling networks



Varun Bhaskar Kothamachu
College of Engineering, Mathematics and Physical Sciences
University of Exeter

A thesis submitted for the degree of Doctor of Philosophy in Mathematics
August 2016

An investigation into dynamic and functional properties of prokaryotic signalling networks

Submitted by **Varun Bhaskar Kothamachu** to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Mathematics.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously submitted and approved for the award of a degree by this or any other university.

Varun Bhaskar Kothamachu

Supervisors

Dr. Ozgur Akman
Prof. Orkun Soyer

Abstract

In this thesis, I investigate dynamic and computational properties of prokaryotic signalling architectures commonly known as the Two Component Signalling networks and phosphorelays. The aim of this study is to understand the information processing capabilities of different prokaryotic signalling architectures by examining the dynamics they exhibit.

I present original investigations into the dynamics of different phosphorelay architectures and identify network architectures that include a commonly found four step phosphorelay architecture with a capacity for tuning its steady state output to implement different signal-response behaviours viz. sigmoidal and hyperbolic response. Biologically, this tuning can be implemented through physiological processes like regulating total protein concentrations (e.g. via transcriptional regulation or feedback), altering reaction rate constants through binding of auxiliary proteins on relay components, or by regulating bi-functional activity in relays which are mediated by bifunctional histidine kinases. This study explores the importance of different biochemical arrangements of signalling networks and their corresponding response dynamics.

Following investigations into the significance of various biochemical reactions and topological variants of a four step relay architecture, I explore the effects of having different types of proteins in signalling networks. I show how multi-domain proteins in a phosphorelay architecture with multiple phosphotransfer steps occurring on the same protein can exhibit multistability through a combination of double negative and positive feedback loops. I derive a minimal multistable (core) architecture and show how component sharing amongst networks containing this multistable core can implement computational logic (like AND, OR and ADDER functions) that allows cells to integrate multiple inputs and compute an appropriate response.

I examine the genomic distribution of single and multi domain kinases and annotate their partner response regulator proteins across prokaryotic genomes to find the

biological significance of dynamics that these networks embed and the processes they regulate in a cell. I extract data from a prokaryotic two component protein database and take a sequence based functional annotation approach to identify the process, function and localisation of different response regulators as signalling partners in these networks.

In summary, work presented in this thesis explores the dynamic and computational properties of different prokaryotic signalling networks and uses them to draw an insight into the biological significance of multidomain sensor kinases in living cells. The thesis concludes with a discussion on how this understanding of the dynamic and computational properties of prokaryotic signalling networks can be used to design synthetic circuits involving different proteins comprising two component and phosphorelay architectures.

To all those who struggle with the detail in their search for the big picture

Acknowledgements

There are several people who need to be acknowledged and thanked for their support and encouragement.

1. Firstly, I would like to thank Prof. Orkun Soyer for his support and advice throughout my PhD. Thank you for supervising me on this project. I have learnt a lot and I hope to imbibe these learnings to become a better researcher. I also thank you for your patience, and positive words during the tough times.

2. I thank Dr. Ozgur Akman, for all his help and support during the last year of my PhD and thesis writing. Dr. Akman, you have been a great support and I really value your advice on my thesis.

3. I thank Prof. Luca Cardelli for a wonderful collaboration, support and advice. Thank you for hosting me at your offices in Cambridge. I treasure the experience, and the learning I gained from our discussions.

4. I would like to specially thank Dr. Elisenda Feliu and Prof. Carsten Wiuf at the University of Copenhagen, for all the advice, learning and a wonderful collaboration. Thank you for hosting me in Copenhagen. It was a very important learning experience for me, and I greatly value the time spent with you.

5. I thank EPSRC and Microsoft Research for funding and support as part of the Dorothy Hodgkin Postgraduate Award and the Microsoft Research PhD programme.

6. I would like to thank the entire group at Prof. O.S. Soyer's lab in Exeter: Arno, Francesco, Deniz, Song, Munia, Jamie, Kalesh and Tobi for their invaluable support, advice and camaraderie when sharing the office. I will always cherish our cake filled group meetings and running the group's coffee machine with Francesco.

7. I would like to thank everyone in Ozgur's group: Jason, Lefteris, Ben for our discussions in the office . It was great sharing an office with you all.

8. I thank Dr Steve Porter for his advice on research questions and collaboration.

9. I would like to thank Phillipe and Barakat for their help and sharing data from P2CS.

10. I would like to thank Dr Chris Ferro, for being an amazing mentor who always helped when I looked to him for advice. Thank you for introducing me to the Erratics Cricket Club and giving me wonderful cricketing memories.

11. I thank all the PhD students, and staff at CEMPS. I would specially like to mention Liz Roberts and Karen Pope from the administrative and finance departments for their support, wonderful banter and all the things they do to support PhD students in the department.

12. I would like to thank all my friends in Exeter: Sneha, Lucky, Bharani, Wessel, Suhaib, Salma, Chinna, and the rest of the lot who have been there for me as friends, inspiring, advising and supporting me during the PhD with wonderful stories, home cooked food and company. My life in Exeter was great because of you all.

13. I thank my parents and my brother who always supported me through the long hours of work and the distance. I could not have done my PhD without you all supporting me.

14. I thank my soon to be wife, Shreya for being a pillar of support, patience and all her love during long hours of writing.

Declaration

The author declares that, to the best of his knowledge, the data contained within this thesis is original and his own work carried out under the supervision of Prof. Orkun Soyer and Dr. Ozgur Akman. The material in this thesis is submitted for the degree of PhD to the University of Exeter only and has not been submitted to any other university. All sources of information have been specifically acknowledged in the form of references.

Contribution Statements

Chapter 2

As per Section 2.2 of code of practice for presentation of Theses, the author states that large portions of chapter 2 are presented as is in the published paper:

Kothamachu VB, Feliu E, Wiuf C, Cardelli L, Soyer OS (2013) Phosphorelays Provide Tunable Signal Processing Capabilities for the Cell. PLoS Comput Biol 9(11):e1003322.

Work done in this chapter was based on previous research on response dynamics in four protein phosphorelays carried out by Prof. Orkun Soyer and Prof. Luca Cardelli. My aim in in this chapter was to thoroughly examine the significance of different biochemical reactions occurring in phosphorelays and examine the effect of these reactions on the overall response dynamics exhibited by these networks. By running numerical simulations using ODE models, extracting published reaction rates and concentrations of phosphorelay proteins, and implementing a multi-parameter sensitivity analysis, I developed an initial theory explaining how different relationships between reaction rates and protein concentrations affect a network's ability to exhibit either sigmoidal, hyperbolic or both sigmoidal and hyperbolic responses at steady state. A subsequent collaboration with Dr. Elisenda Feliu and Prof. Carsten Wiuf lead to the analytical solution for ODEs representing a phosphorelay, and helped validate my initial observations on tunable phosphorelays and the relationships between parameters in the network which give rise to different response dynamics.

When comparing the reliability of responses from tunable phosphorelays exhibiting both sigmoidal and hyperbolic regimes, I used PRISM, a probabilistic model checker tool introduced by Prof. Cardelli, writing PRISM scripts representing a tunable four protein phosphorelay architecture and comparing the variance in response from a phosphorelay in sigmoidal and hyperbolic response regimes. In addition to the

above, I carried out all other remaining analyses discussed in this chapter, created the figures and generated all data shown in all the tables. Additional work done by Dr. Elisenda Feliu and Prof. Carsten Wiuf has been included in the appendices as they are necessary to understand some of the derivations in the main text.

Chapter 3

As per Section 2.2 of code of practice for presentation of Theses, the author states that large portions of chapter 3 are presented as is in the paper:

Kothamachu VB, Feliu E, Cardelli L, Soyer OS (2015) Unlimited multistability and Boolean logic in microbial signalling. Journal of The Royal Society Interface 12(108):20150234.

In this chapter, I study the effects of different sensor kinases i.e. regular and multidomain HKs on the stability and signal processing capabilities of a phosphorelay. Using the CRNT toolbox and analysing the chemical reaction network underlying hybrid HK containing phosphorelays, I found that phosphorelays containing multidomain sensor HKs (i.e. hybrid and unorthodox) have the capacity for multistability. By examining their network architecture, I identified a minimal core sub-network which is bistable. From this, I derived an initial explanation for bistability in the core network occurring due to coupled feedback loops. Using ODEs representing reactions in this network, I derived an analytical solution for the system and identified necessary conditions for bistability in the core network. Dr. Elisenda Feliu contributed towards generalising these findings from the core bistable network to the full phosphorelay network containing a hybrid HK and an unorthodox HK.

Building on this, I started exploring the behaviour of networks containing these multidomain HKs and sharing proteins (components) to integrate multiple signals and response. My initial derivation for a system with a two multistable cores sharing a single Hpt ($n = 1$), showed that the system had $2n + 1$ steady states. This was expanded and generalised by Dr. Elisenda Feliu to explain the behaviour of systems with $n \geq 2$ multistable cores sharing a common Hpt and then a common RR.

In addition to the correlation of multistability with the number of pathways sharing proteins, I started examining the capacity of these architectures to implement different logical operations in the cell. Using ODEs representing different architectures with shared RR and Hpt, and writing scripts running numerical simulations that sampled different parameter sets, I identified parameters where some networks with shared

proteins implemented AND, OR and ADDER operations in the cell.

Contents

List of Figures

List of Tables

List of Symbols

ATP Adenosine Tri Phosphate

ADP Adenosine Di Phosphate

BLAST Basic Local Alignment Search Tool

CTMC Continuous Time Markov Chain

DAG Directed acyclic graph

DNA Deoxyribonucleic Acid

EXP Inferred from Experiment

FDR False Discovery Rate

FWER Family wise Error Rate

GI ID Represents sequence records processed by NCBI.

GO IDs Gene Ontology IDs

HK Histidine Kinase protein

HMM Hidden Markov Models

HPT Phosphotransfer protein

nr Non redundant database

NCBI National Center for Biotechnology information

MAPK Mitogen-activated protein kinases

ODE Ordinary Differential Equations

PIR Protein Information Resource

P2CS Prokaryotic 2- Component Systems database

PRISM Probabilistic Model checker

SNR Signal to Noise Ratio

REC Receiver protein

RR Response Regulator protein

TCS Two Component Signalling

XML Extensible Markup Language

1

Introduction

Decision making in prokaryotic cells occurs through a complex web of signalling networks embedding both information gatherers and signal transducers. They help the cell to sense changes in its external and internal environments and respond to it by modulating various physiological processes in the cell. The term architecture represents the structure of a signalling network defined by its constituent proteins and reactions occurring amongst these proteins when a signal molecule flows through the network. As these signal molecules travel from one member of the network to another, they carry information about the environment from the information gathering sensor to the transducer (response regulator) protein that modulates a response from the network. These networks are therefore important for cells to capture and process information about their environment, and ensure that the cells deal with changes (around and within) in a reliable and timely manner.

One of the simplest prokaryotic signalling architectures has only two proteins, namely the sensor and transducer proteins. This forms a simple signalling network known as the two component signalling (TCS) network (see figure 1.1A). Variants of this architecture containing additional proteins also occur in prokaryotic cells, and are known as phosphorelays. A schematic representation of the simple TCS and a phosphorelay network is shown in figure 1.1. Unlike a simple TCS, phosphorelays

have two additional phosphotransfer steps involving a signal molecule as it moves from the sensor to its cognate response regulator. These additional proteins (as shown in figure 1.1B) are known as the receiver (REC) and phosphotransfer (Hpt) proteins.

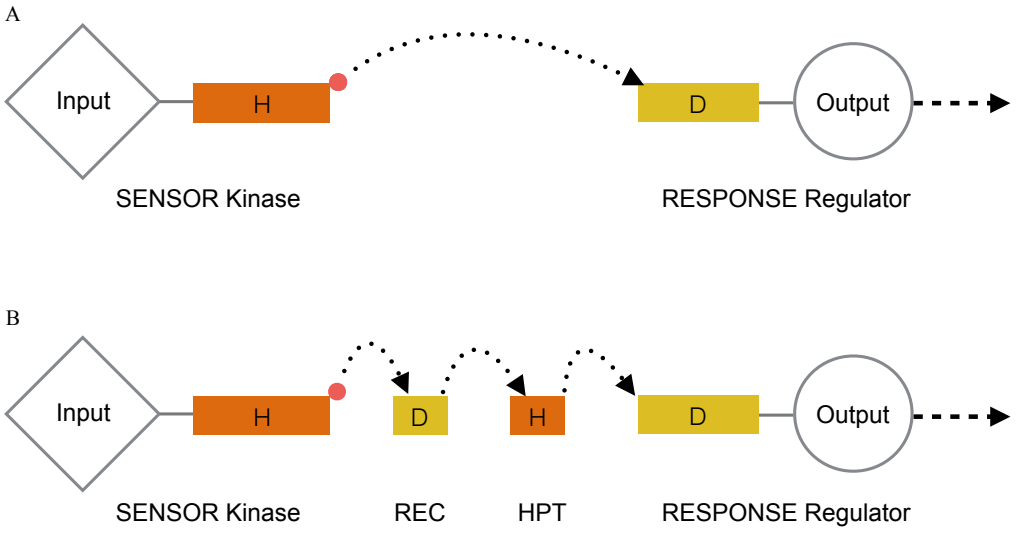


Figure 1.1: Shows a two component signalling architecture (A) and phosphorelay (B). The red filled circle represents phosphate groups and arrows show the direction in which phosphate groups are transferred.

With numerous physiological processes in a cell, all occurring over a range of time scales and requiring different dynamic behaviours, some prokaryotic cells like *Pseudomonas aeruginosa* employ hundreds of variants of simple TCS and larger phosphorelay networks [1]. Networks like ArcB-ArcA [2] forms a simple two component signalling network that regulates gene expression in a cell, depending on the presence (or absence) of oxygen; the KinA-Spo0A pathway found in *Bacillus subtilis* has four proteins which form a phosphorelay architecture that regulates sporulation [3]. These signalling networks in the cell can either operate individually, regulating disparate processes, or in tandem [4] where they communicate with each other by sharing proteins, use connector molecules that amplify signals [5]; or via crosstalk [6] where signal molecules from one network travels to another. Even as signalling between members of a signalling network is highly specific [7], some networks employ crosstalk (phosphorelays) and embed interesting dynamics [6, 8, 9] while processing signals in the cell. This presence of a number of signalling architectures and communications between them show the complex signal processing and regulatory strategies occurring within prokaryotic cells.

While these networks operate within the cell, it is important that they map different signals, or combinations thereof, onto specific physiological responses within the cell in a responsive and reliable fashion. To understand the mechanisms that define signalling architectures and how they ensure a reliable and responsive mapping of signals, this study focusses on commonly found prokaryotic signalling architectures and examines the significance of different biochemical reactions and the resultant architectures involved in prokaryotic signalling. Using methods borrowed from dynamical systems theory, this study uses differential equation models representing reactions in these networks, and studies their dynamic and computational properties. Studying different architectures, and understanding the significance of different biochemical reactions and proteins occurring in signalling networks can explain how modular prokaryotic signalling architectures are capable of exhibiting complex behaviours like adaptation (seen in chemotaxis[10]) or a food per cell computation carried out when regulating sporulation [3] in *Bacillus subtilis*.

In comparison to eukaryotic signalling, prokaryotes have simpler signalling architectures.[1] Understanding these networks could help us derive an insight into the evolution of complexity in cells and the dynamic and computational properties exhibited by networks containing different types of proteins could highlight the evolutionary context for proteins which play the role of information gatherers and transducers in the cell. Understanding this is important as these proteins are required in a cell to constantly sense changes in the cell's environment and respond to it by modulating its internal cellular machinery.

This study does not directly examine evolutionary mechanisms underlying prokaryotic signalling networks but focuses instead on characterising the functioning of different signalling architectures and their dynamic and computational properties. This is to identify possible design principles that define the structure of signalling architectures based on the signal processing capabilities required by the cell to continuously track changes in its environment and respond to them reliably. Building on initial investigations from a system dynamics perspective, this study attempts to find the biological significance of different signal processing characteristics exhibited by prokaryotic signalling networks. The next section gives a brief background into the biochemistry of two component and phosphorelay signalling.

[1]Unlike prokaryotes, eukaryotes are multicellular organisms with a more complex structure that includes membrane bound organelles and a nucleus with a nuclear membrane.

1.1 Introduction to Signalling

1.1.1 Two Component Signalling networks

As discussed at the beginning of this chapter, the main components of a simple two component signalling architecture are its sensor and response regulator proteins. The sensor protein usually has a kinase activity, where it acts as an enzyme that transfers a phosphate group from a phosphate containing molecule to an unphosphorylated substrate. The sensor kinase is usually a membrane bound receptor protein with an extracellular input domain[2] that detects specific signalling molecules in its environment, changes its active conformation [11] and undergoes an ATP[3] mediated autophosphorylation reaction to receive a γ phosphate group from the ATP. The autophosphorylated sensor kinase protein then transfers the phosphate group onto an aspartate containing receiver portion (also called domain) of the response regulator (RR) found in the cytoplasm of the cell. The RR acts as a transducer which receives information from the sensor kinase (through phosphotransfer reactions) and regulates the output of downstream process in the cell.

Reactions in a two component signalling network

$$\mathrm{HK} \xrightarrow{k_s} \mathrm{HK_p} \tag{1}$$

$$\mathrm{RR} + \mathrm{HK_p} \underset{k_r}{\overset{k_1}{\rightleftharpoons}} \mathrm{RR_p} + \mathrm{HK} \tag{2}$$

$$\mathrm{RR_p} \xrightarrow{k_2} \mathrm{RR} \tag{3}$$

Once the response regulator receives a phosphate group from a sensor kinase, the phosphorylated RR can play a variety of downstream roles [12, 13] depending on the processes that it is involved in. Upon receiving phosphate groups, the RR either activates or deactivates a downstream signalling pathway [14], acts as a transcription factor which directly binds to the cell's DNA and regulates gene expression [15] or acts as a sink protein that creates a tunable threshold device [16]. Once a phosphorylated response regulator loses its phosphate group after affecting a downstream process, the signal is lost from the network and the network's response is

[2] A domain in a protein is a portion of the protein which is independently stable, and can maintain its structure and function, even in the absence of the rest of the protein.

[3] ATP stands for Adenosine Tri Phosphate. It is a phosphorylated nucleotide (adenosine) with three phosphate groups, α, β, γ. ATP is a common source of energy in various biochemical processes within the cell.

terminated. This flow of a phosphate group from a sensor to the response regulator represents the flow of information through in a simple prokaryotic signalling network.

Histidine kinases are one of the most commonly found sensor kinases in prokaryotes and occur occasionally in eukaryotes. In general, there are a several types of sensor kinases defined by the amino acids found on their phosphorylation site. If it is a histidine residue, the sensor protein is called a histidine kinase. If there are other amino acids like serine, threonine and tyrosine kinases, the kinases are named accordingly [11] as serine, threonine or tyrosine kinase. A preference for histidine kinase in prokaryotes could be due to the formation of a phosphoramidate bond when a histidine kinase binds with an incoming phosphate group. This bond has a high free energy and low stability which allows cells to break this bond easily during signalling, and makes the response time for the transfer of signal molecule from the histidine kinase very short.

When sensor kinases are only involved in the transfer of phosphate groups onto its substrate or the next member in the relay, they are known as a monofunctional kinases. There are some systems [17–22], where the kinase protein performs an additional role of removing phosphate groups from a protein i.e. phosphatase activity, they are known as a bifunctional sensor kinase. This dual role for a sensor kinase allows a signalling network with a bifunctional histidine kinase to respond to specific triggers, or stress molecules, which bind to the sensor kinase and toggle the sensor protein's activity from a kinase to a phosphatase or vice versa [23]. This can be a useful mechanism employed by signalling networks to tune signal response behaviours in some architectures or as a means of building a buffer against crosstalk in the system [24]. When carrying out a phosphatase activity, the sensor kinase forms a complex with an already phosphorylated aspartate domain in the relay which, in some cases can also lead to interesting dynamics [25] due to a temporary sequestration of a downstream signalling member [26] in the relay.

In summary, the basic two component signalling architecture with a sensor histidine kinase and response regulator represents one of the simplest signalling architectures in prokaryotes. It is commonly found across bacteria [1], archaea [27] and in some eukaryotes [28] also. In some cases, the modularity of this simple architecture is leveraged by a cell to evolve longer variants, i.e. phosphorelays. The following section discusses phosphorelay architecture in more detail.

1.1.2 Phosphorelays

As described earlier (and shown in figure 1.1B), a phosphorelay is an extension of a simple two component signalling architecture and contains additional receiver and phosphotransfer domains involved in phosphotransfer within the network. These domains can either occur on separate proteins (figure 1.1B) or on the same protein as shown in figure 1.2. Sensor kinases with an additional REC domain along with a kinase are known as a hybrid sensor kinase (see figure 1.2A) and while those with have both REC and Hpt domains are known as unorthodox sensor kinase proteins (figure 1.2B).

In phosphorelays, the signal molecule (phosphate group) flows from a sensor histidine kinase to a receiver protein and reaches Hpt before being transferred on to a response regulator (See figure 1.2A). In a number of systems, phosphotransfer reactions between REC-Hpt and Hpt-RR are reversible [29, 30]. The flow of phosphate groups amongst domains found in phosphorelays, is always from a histidine to aspartate or *vice versa*. There is no phosphotransfer between similar domains (i.e. histidine→ histidine or aspartate→ aspartate). This conserved nature of phosphotransfer is shared by all prokaryotic signalling architectures which have these proteins [31] and is used to build modular signalling architectures in a cell.

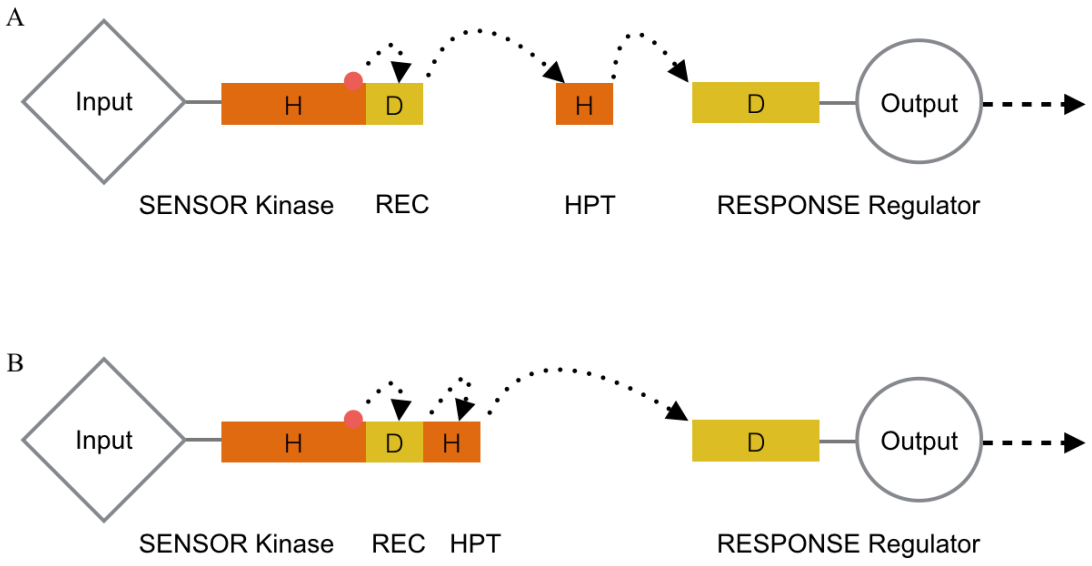


Figure 1.2: Shows a Hybrid Sensor Kinase (A) and an Unorthodox Sensor Kinase (B).

Reactions in a phosphorelay

$$\mathrm{HK} \xrightarrow{k_s} \mathrm{HK_p}$$

$$\mathrm{HK_p} + \mathrm{REC} \xrightarrow{k_2} \mathrm{HK} + \mathrm{REC_p}$$

$$\mathrm{REC_p} + \mathrm{Hpt} \xrightarrow{k_3} \mathrm{REC} + \mathrm{Hpt_p}$$

$$\mathrm{Hpt_p} + \mathrm{RR} \xrightarrow{k_4} \mathrm{Hpt} + \mathrm{RR_p}.$$

$$\mathrm{HK} + \mathrm{REC_p} \xrightarrow{k_{2r}} \mathrm{HK_p} + \mathrm{REC}$$

$$\mathrm{REC_p} \xrightarrow{k_{h1}} \mathrm{REC}$$

$$\mathrm{REC} + \mathrm{Hpt_p} \xrightarrow{k_{3r}} \mathrm{REC_p} + \mathrm{Hpt}$$

$$\mathrm{RR_p} \xrightarrow{k_{h2}} \mathrm{RR}$$

$$\mathrm{Hpt} + \mathrm{RR_p} \xrightarrow{k_{4r}} \mathrm{Hpt_p} + \mathrm{RR}.$$

A widely held view is that phosphorelays have evolved to allow the cell to achieve signal integration using phosphorylation at their different layers[4] [30, 32]. While theoretical studies have shown the potential of relays for signal integration [3, 33], it remains unclear how and why selective pressures on signal integration should lead to the widespread phosphorelay features such as relay length of four and presence of reverse phosphorylation. Some investigations into the length (number of phosphotransfer steps) of a phosphorelay show that phosphorelays with four proteins achieve an optimal sensitivity to incoming signals and having any additional proteins (> 4) in the network brings no significant improvement in its ability to sense signals [33] from the environment. This could explain the absence of naturally occurring phosphorelays with more than four proteins in a single phosphorelay network but it remains to be proven conclusively.

An alternative possibility is that evolution of phosphorelays has resulted in specific features due to their effects on signal processing capabilities of the cell. For example, transcriptional feedbacks such as those seen in the *Bacillus subtilis* phosphorelay raise the possibility of achieving bistable dynamics as a functional role of the relay

[4]Layers refer to proteins in the relay. Both these terms mean the same and are used interchangeably in this thesis.

[34, 35]. Recent experimental studies, however, show that this system possesses no bistability and that the embedded feedback loops might have limited effect on its steady state response [36, 37]. One of these studies found high heterogeneity in the phosphorylated RR levels [36], while the other showed experimental evidence for ultrasensitivity in the output of the phosphorelay but suggested that the source of this feature lies downstream of the relay [37]. The former result leads to the proposal that the main functional role of the *B. subtilis* sporulation phosphorelay is to act as a noise generator. It is not clear how these findings and hypotheses should apply to other phosphorelays, and particularly to those that are not nested in transcriptional feedbacks. Knowing that cells can employ different signalling arrangements (topologies[5]) and sometimes use multiple networks to integrate multiple signals and regulate responses in multiple pathways via branched signalling networks, this study examines branched signalling networks with shared proteins and signalling molecules along with individual two component and phosphorelay architectures.

Functional arguments derived from studies of a specific system are usually limited in providing a broad understanding of the relation between the observed architectural and biochemical features of phosphorelays and their function. This study therefore carries out a detailed mathematical investigation into different variants of prokaryotic signalling architectures and develops a thorough understanding of their functioning in the cell.

1.2 Introduction to Modelling Approaches

Different signalling networks exhibit varying dynamics like oscillations [38], hyperbolic [39]), sigmoidal [40] or show a rapid adaptation [41] to changes in the external environment of the network. Understanding the mechanisms causing these dynamics is essential to identify potential design principles dictated by the correlation between dynamics of different signalling architectures and the biological functions they regulate. The dynamics of signalling networks involving a series of biochemical reactions with a complex interplay and involvement of biomolecules from different cellular processes are usually modelled using a system of ordinary differential equations (ODE). This section therefore introduces ordinary differential equations and describes how we express chemical reactions in systems using ODEs. We take the example of a simple two component signalling network (see figure 1.1) and discuss the assumptions and simplifications made when using mass action

[5]We use the terms *topology*, *architecture* and *arrangement* interchangeably in this thesis

kinetics to derive these ODEs. As discussed in the previous section, a two component signalling network has a commonly found prokaryotic signalling architecture and is found across a wide number of genomes [11, 42, 43].

Dynamical systems theory is an area of mathematics used to describe complex temporal dynamics with models containing either difference or differential equations [44]. For systems where the change in time is continuous (like signalling networks), we use differential equations and for those where the changes are discrete, difference equations are more appropriate. In this section, we describe continuous time dynamical systems and discuss how differential equations are used to understand the temporal dynamic behaviours exhibited by two component signalling networks. We begin with an introduction to differential equations and show how an ordinary differential equation model for a prokaryotic two component signalling network can be derived.

1.2.1 Ordinary Differential Equations

Differential equations represent the evolution of one or more (dependent) variables with respect to one or more (independent) variables by specifying mathematical functions that relate the derivatives of the dependent variables to the values of the independent variables. If the derivatives are calculated with respect to a single independent variable (for example, equation (4)), then the equation is known as an ordinary differential equation (or ODE) but if the derivatives are calculated with respect to more than one independent variable, then the equation is called a partial differential equation (PDE) (for example, equation (5)).

$$\frac{dx}{dt} = a \cdot x + b \tag{4}$$

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{1}{v^2} \cdot \frac{\partial^2 u}{\partial t^2} \tag{5}$$

We know that two component signalling networks have two key members viz. the sensor protein (HK) and the response regulator (RR). Let us consider reactions (1)-(3) occurring in a simple two component signalling network. This network has the two proteins HK and RR occurring in either their phosphorylated (HK_p, RR_p) or unphosphorylated forms (HK, RR). These phosphotransfer reactions in two component systems are not enzyme mediated and occur when a phosphate carrying member of the network collides with a cognate unphosphorylated signalling partner.

In reaction (1), we see that HK is phosphorylated at a rate k_s to form HK_p. We are assuming that there is sufficient ATP in the surrounding environment that the reaction rate is independent of changes in ATP concentration, reducing this phosphotransfer reaction to a unimolecular reaction. The other phosphotransfer reactions in the network occur at rates k_1 and k_2, representing the phosphotransfer from HK_p to RR and RR_p losing the phosphate group through a dephosphorylation or an auto-hydrolysis reaction (3). The rate k_r represents the rate at which a reverse phosphotransfer where the phosphate group from RR_p travels back to a molecule of HK. In this model, auto-hydrolysis shown in reaction reaction (3) also accounts for the loss of phosphate groups from RR due to a response triggered by its action on a downstream process.

When modelling these reactions using ordinary differential equations, we assume mass action kinetics [45], and consider the stoichiometry[6] of the reactions (1)-(3) to determine the coefficients of equations (6)-(9) representing changes in the concentration of proteins in this network. When using mass action kinetics, we make the assumption that these chemical reactions occur in a well mixed system at equilibrium and the rate of the reaction is directly proportional to the concentration of reactants and the probability (usually called reaction rate) of collisions between them. This means that, in reaction (2), the rate of formation of RR_p is directly proportional to the product of the concentration of both the reactants HK_p and RR, and the probability of collisions between the two (i.e reaction rate k_1). In general, for any protein P in this network, concentration at any given point of time ($\frac{dP}{dt}$) is the difference between reactions producing new molecules of P and those removing P from the system. If we consider the example of HK in the system, new molecules of HK are created in the forward phosphotransfer reaction (2) and removed in the reverse phosphotransfer reaction (2) and in reaction (1). The ODE describing the rate of change of HK (i.e. $\frac{dHK}{dt}$) can be written as shown below.

$$\frac{dHK}{dt} = k_1 \cdot RR \cdot HK_p - k_s \cdot HK - k_r \cdot HK \cdot RR_p \tag{6}$$

We repeat this approach for every protein in the network and derive a system of ODEs which represent the change in concentration of all phosphorylated and unphosphorylated proteins in a two component signalling network (6)-(9).

[6]Stoichiometry of chemical reactions represents changes in the number of molecules of all species found in a reaction.

$$\frac{dHK_p}{dt} = k_s \cdot HK + k_r \cdot RR_p \cdot HK - k_1 \cdot RR \cdot HK_p \tag{7}$$

$$\frac{dRR}{dt} = k_r \cdot RR_p \cdot HK + k_2 \cdot RR_p - k_1 \cdot RR \cdot HK_p \tag{8}$$

$$\frac{dRR_p}{dt} = k_1 \cdot RR \cdot HK_p - k_r \cdot RR_p \cdot HK - k_2 \cdot RR_p \tag{9}$$

Building this ODE model is useful for understanding how concentration of different proteins in a dynamic system (like signalling network) changes over time. In addition to tracing the temporal variation in the variables of the system, ODE models allow us to identify points in the system's evolution where the net change in the concentration of different variables of the system over time becomes zero, and the system reaches its steady state. By equating the right hand side (R.H.S) of ODEs in expressions (6-9) to zero, we can derive the steady state solution of a two component signalling network as a function of $\mathrm{RR_p}$ and the reaction parameters defining the rate at which different phosphotransfer reactions in the network take place.

In the following sections, we discuss methods used to study the temporal evolution and steady state behaviour of a system of differential equations defining two component signalling networks.

1.2.2 Numerical Simulations

When a dynamical system starts with some initial conditions, depending on the initial values of different variables in the system, and the equations defined as ODEs, numerical methods like the Euler method [44, 46], mid point method [44, 46, 47] and Runge-Kutta (`http://mathworld.wolfram.com/Runge-KuttaMethod.html`) [47, 48] method can be used to iteratively derive the approximate values for different variables in the system at different time steps. This approach of iteratively solving a set of ODE equations using time steps is referred to as numerical simulations.

The accuracy of approximating values of different variables using methods like Euler or Runge-Kutta depends on the size of the time step dt as a smaller time step indicates a finer mesh which accounts for both fast and the slow reactions in a system. A smaller time step increases the accuracy of the approximation, but at the same time, it increases computational time and hence the resources required to numerically simulate the system increases very quickly. This is more

Parameters	Low	High	Source
k_2 $(Ms)^{-1}$	10	10^5	[54]; [55]; [24]
k_3 $(Ms)^{-1}$	10	10^5	[54];[55]; [24]
k_4 $(Ms)^{-1}$	10	10^5	
k_5 $(Ms)^{-1}$	10	10^5	[22] ; [54]
k_6 $(Ms)^{-1}$	1	10	[22] ; [54]
k_{2r} $(Ms)^{-1}$	10	10^5	[54]
k_{3r} $(Ms)^{-1}$	10	10^5	[54]
$k_{4r}(Ms)^{-1}$	10	10^5	[54]
k_{5r} $(Ms)^{-1}$	10^{-3}	10^{-1}	[22] ; [54]
k_{h1} $(s)^{-1}$	10^{-3}	10^{-1}	[54] ;[55]
k_{h2} $(s)^{-1}$	10^{-3}	10^{-1}	[54] ;[55]
HK_{total} (M)	10^{-6}	2 x 10^{-4}	[54]
REC_{total} (M)	10^{-6}	2 x 10^{-4}	[54]
Hpt_{total} (M)	10^{-6}	2 x 10^{-4}	[54]
RR_{total} (M)	10^{-6}	2 x 10^{-4}	[54]

Table 1.1: Shows parameter regimes for reaction rates and protein concentrations derived from published experimental data. Sources for these parameters are listed in the fourth column.

challenging with stiff systems, where the time scales of the slowest and quickest reactions differ significantly, and therefore numerical simulations become more unstable.

In this study, we carry out a detailed investigation on the role of all the different parameters in prokaryotic signalling networks and their effects on prescribing specific dynamics to the network. Table 1.1 shows biological parameters from published experimental data, used in these numerical simulations. We encode reactions in the network as ODE functions on Matlab [49] and use solvers like ode45s, ode23t and ode15s [50] to iteratively solve the system of equations represented in the ODE function. ode45 and ode23 use Runge-Kutta-Fehlberg integration algorithm, with a variable step size while ode15s is a variable order solver which uses numerical differentiation formulas [50, 51]. In addition to using Matlab, tools like XPPAUT (`http://www.math.pitt.edu/~bard/xpp/xpp.html`) [52] and Oscill8 [53] can also be used for numerically simulating a system of ODEs.

In addition to understanding the temporal dynamics of signalling networks, it is also important to know how the system behaves when it reaches a steady state. For this, we can derive an analytical solution that gives us expressions for the steady state values of different variables in the network. This involves solving for each variable from the ODE model in a logical manner where the final outcome is a series of expressions for each variable in the system as a function of other variables and reaction parameters. The steps involved in deriving the steady state analytical solution for a system of ODEs representing two component signalling networks are described in section 1.2.3. When dealing with large and complex system of ODEs, it is sometimes difficult to arrive at an analytical solution for the system. In this case, numerical simulations are useful in approximating the concentration of different proteins in the network over time and calculating the approximate steady state value of different variables in the system by ensuring that they have converged to a value which does not vary, even if the system is numerically simulated for an additional time period.

1.2.3 Steady State Analytical Solution

As mentioned in the previous section, a key aspect of understanding a network's dynamics is knowing how the system behaves when it is at steady state. We know that mathematically, a steady state of a system is a point in variable space at which the variables do not change with time. In some cases, the steady state of a system is stable and in others, it is unstable. When we identify that a system has a stable steady state, it means that a small perturbation at the steady state will damp out, resulting in the system settling back into its original steady state (figure 1.3A). But in the case of an unstable steady state, fluctuations are amplified and instead of being dampened, they push the system away from the original steady state (figure 1.3B). To derive such an understanding of the steady state of a system, we need to examine the sign of the real parts of the eigenvalues of the system obtained by linearising the equations evaluated at steady state. Negative real parts indicate stability, whilst one or more positive parts indicate instability [56].

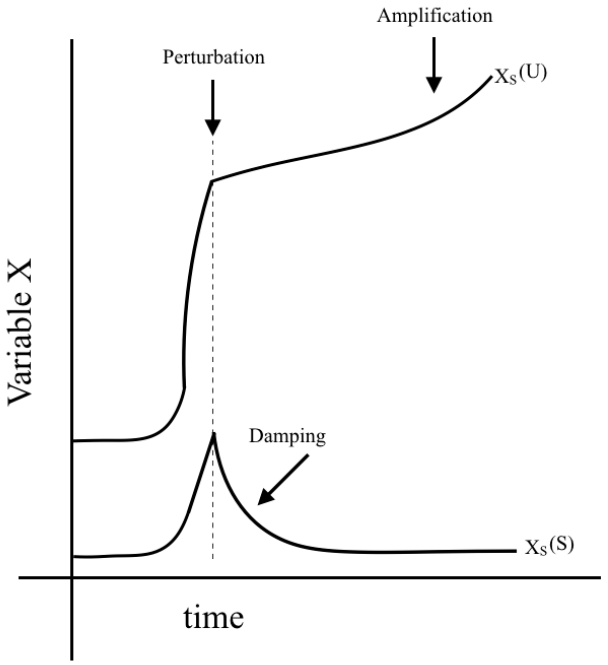


Figure 1.3: Shows two types of stability i.e. stable (X_S(S)) and unstable (X_S(U)) steady state.

In this study, we derive analytical solutions for ODEs (6)-(9) by equating their right hand side (R.H.S) to zero. Solving the resultant system of equations gives the value of different variables in the system at steady state. This is known as an analytical or exact solution for the system. In the context of a biological signalling network which regulates gene expression (for example [57]), the steady state concentration of proteins controlling gene expression in the system defines different expression patterns in the cell,and can result in different physiological states of a cell. In (6)-(9), when we equate the right hand side of these expressions to zero, it means that $\frac{dHK}{dt} = \frac{dHK_p}{dt} = \frac{dRR}{dt} = \frac{dRR_p}{dt} = 0$, we get

$$0 = k_1 \cdot RR \cdot HK_p - k_s \cdot HK - k_r \cdot HK \cdot RR_p \tag{10}$$

$$0 = k_s \cdot HK + k_r \cdot RR_p \cdot HK - k_1 \cdot RR \cdot HK_p \tag{11}$$

$$0 = k_r \cdot RR_p \cdot HK + k_2 \cdot RR_p - k_1 \cdot RR \cdot HK_p \tag{12}$$

$$0 = k_1 \cdot RR \cdot HK_p - k_r \cdot RR_p \cdot HK - k_2 \cdot RR_p \tag{13}$$

We see that (10) = −(11) and (12) = −(13). These equations are redundant and we therefore reduce the number of equations by removing (10) and (12). From the resultant ODEs, we find the total concentration of both phosphorylated and unphosphorylated forms of all proteins in the network is conserved i.e. $HK + HK_p = HK_{tot}$ and $RR + RR_p = RR_{tot}$. We can therefore rewrite equations (11) and (13) to arrive at smaller system of ODEs where all variables in the system are expressed in their phosphorylated forms (HK_p and RR_p) as shown below.

$$0 = k_s \cdot (HK_{tot} - HK_p) + k_r \cdot RR_p \cdot (HK_{tot} - HK_p) - k_1 \cdot (RR_{tot} - RR_p) \cdot HK_p \tag{14}$$

$$0 = k_1 \cdot (RR_{tot} - RR_p) \cdot HK_p - k_r \cdot RR_p \cdot (HK_{tot} - HK_p) - k_2 \cdot RR_p \tag{15}$$

Expanding and rearranging equations (14) and (15) gives us,

$$0 = ((k_1 - k_r) \cdot RR_p - k_1 \cdot RR_{tot} - k_s) \cdot HK_p + k_r \cdot RR_p \cdot HK_{tot} + k_s \cdot HK_{tot} \tag{16}$$

$$0 = ((-k_1 + k_r) \cdot RR_p + k_1 \cdot RR_{tot}) \cdot HK_p + (-HK_{tot} \cdot k_r - k_2) \cdot RR_p \tag{17}$$

By fixing the values of RR_p, i.e. assuming that we know its values at steady state, we use a recursive approach to derive an expression for all the other variables in this system as a function of RR_p. From equation (17), we isolate an expression for HK_p as a function of RRp .

$$HK_p = -\frac{RR_p \cdot (HK_{tot} \cdot k_r + k_2)}{RR_p \cdot k_1 - RR_p \cdot k_r - RR_{tot} \cdot k_1} \tag{18}$$

To derive an expression for HK, we consider the conservation of total HK concentration in the cell, i.e. $HK = HK_{tot} - HK_p$

$$HK = \frac{HK_{tot} \cdot RR_p \cdot k_1 - HK_{tot} \cdot RR_{tot} \cdot k_1 + RR_p \cdot k_2}{RR_p \cdot k_1 - RR_p \cdot k_r - RR_{tot} \cdot k_1} \tag{19}$$

and for RR, we know that

$$RR = RR_{tot} - RR_p \tag{20}$$

For signalling networks like TCS, a steady solution for the system is only valid when all the variables in the system have positive values. Using equation (18) and (19), for any positive value of RR_p, we know that HK and HK_p are positive when

$$RR_p < \frac{RR_{tot} \cdot k_1}{k_1 - k_r} \tag{21}$$

As $RR = RR_{tot} - RR_p$, we know that RR is positive as

$$RR_p < RR_{tot} \tag{22}$$

Expressions (21) and (22), give us the bounds for values of different variables in the system, within which steady state values of all variables in the system remain positive.

1.2.4 Signal Response Relationships

Building on our understanding of the temporal and steady state behaviour of signalling networks, deriving an understanding of how a dynamic system responds to different inputs is essential for understanding the signal processing capabilities of signalling networks. When deriving an expression for the steady state signal-response curve for all the prokaryotic signalling networks considered in this study (analytically), we consider k_s as the input and the concentration of RR_p as a response from the system. The expression for a signal response curve , i.e. k_s vs RR_p is shown below.

$$k_s = f^{-1}(RR_p)$$

By substituting values for HK_p and RR_p in equation (16), we can derive an expression for k_s as

$$k_s = \frac{RR_p \cdot k_2 \cdot (RR_p \cdot k_1 - RR_p \cdot k_r - RR_{tot} \cdot k_1)}{HK_{tot} \cdot RR_p \cdot k_1 - HK_{tot} \cdot RR_{tot} \cdot k_1 + RR_p \cdot k_2} \tag{23}$$

Using this expression, we can generate steady state signal-response curves for a simple two component signalling network (k_s vs RR_p). When we consider steady state signal-response curves, it is important to understand the nature of the signal-response curve and the effects that different parameters in the model have on its shape.

1.2.5 Classifying steady state signal-response curves

The shape of a signal-response curve carries important information about the signal processing capabilities of a system [2]. For example, a curve with a strong sigmoidal shape indicates a switch-like response, in which the system acts like a filter, where signals below a certain threshold do not generate any significant response. In the case of a linear signal-response curve, the system responds to all incoming signals, while a hyperbolic signal-response curve means that the system responds to all input signals until it reaches a saturation point, after which the output remains the same.

The shape of a signal-response curve has been quantified in some systems using the ratio of signals generating 90% and 10% of maximal response [40]; however, measuring the level of sigmoidality for any arbitrary signal-response curve is not trivial. Before the analytical solution was derived, attempts were made to classify steady state signal response curves by fitting signal response curve data with a Hill function and Monod-Wyman-Changeux functions to derive a measure for sigmoidality and then classify steady state signal-response curves into sigmoidal or non sigmoidal curves. These attempts resulted in poor fits and this classification approach was not very reliable when sampling over the wide range of parameters (extracted from experimental data) used in our models.

To get around these challenges of curve fitting and measuring signals generating 90% and 10% of maximal response, scripts that count the number of sign changes in the second derivative of RR_p as a function of k_s in the signal-response plot can be used. For a sigmoidal curve, there is usually a change in sign from positive to negative (see figure 1.4) whilst in the case of a hyperbolic curve, the second derivative has no change in the sign and is always negative. This classification approach was implemented through scripts in Matlab.

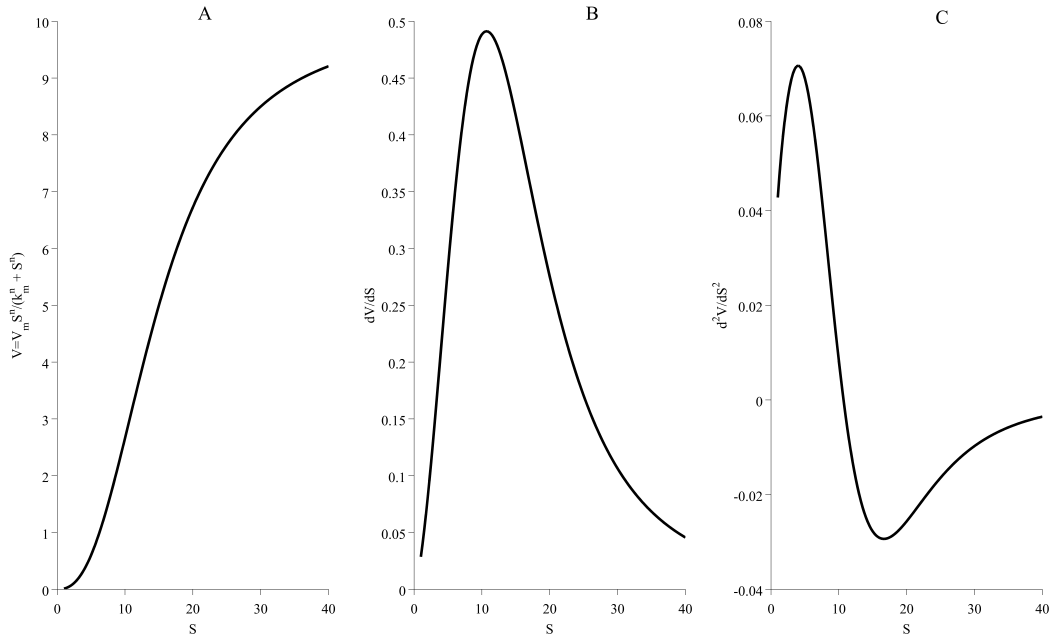


Figure 1.4: Represents a sigmoidal curve derived using a hill equation $V = \frac{V_m \cdot S^n}{k_m^n + S^n}$. S is plotted on the x-axis, and A, B & C show , the corresponding first derivative $\frac{dV}{dS}$ and the second derivative $\frac{d^2V}{dS^2}$ respectively, plotted on the y-axis. Parameters used to generate this plot were $V_m = 10$; $n = 2.5$; $k_m = 15$. The values of S were varied between 0 and 40.

Once we derived the analytical expression for the k_s vs RR_p curve, we classified curves easily by examining the second derivative of the equation (23), representing the steady state signal-response curve. We checked if the slope of the second derivative curve is positive or negative at $RR_p = 0$. If the curve is sigmoidal, the second derivative curve is initially positive and then becomes negative (see figure 1.4) while hyperbolic curves will always have a negative second derivative. Thus, the sign of the second derivative of the signal-response curve at zero can be taken as a test for sigmoidality. This was also confirmed by the agreement of sigmoidality classifications based on the sign of the second derivative of the signal-response curve on the entire curve. Using this criterion, we grouped the curves into sigmoidal or hyperbolic curves.

From (23), we derive the second derivative expression for k_s as a function of RR_p.

$$k_s'' = -\frac{2 \cdot k_2 \cdot HK_{tot} \cdot RR_{tot}^2 \cdot k_1^2 \cdot (HK_{tot} \cdot k_r + k_2)}{(HK_{tot} \cdot RR_p \cdot k_1 - HK_{tot} \cdot RR_{tot} \cdot k_1 + RR_p \cdot k_2)^3} \tag{24}$$

At $RR_p = 0$,we find that

$$k_s''(0) = \frac{2 \cdot k_2 \cdot (HK_{tot} \cdot k_r + k_2)}{(k_1 \cdot RR_{tot} \cdot HK_{tot}^2)} \tag{25}$$

Expression (25) is always positive and this shows that a simple TCS network with separate sensor HK and response regulator RR exhibits a sigmoidal signal response behaviour.

1.2.6 Detecting Multistability

When we examine the steady state behaviour of different dynamical systems, we find that some systems can have more than one steady state [58–62]. If a system has a single stable steady state, irrespective of what its initial conditions are, the system will always converge to the same value. By contrast, in multistable systems, depending on the initial conditions, the final steady state values of the variables can be different. Figure 1.5A shows an example schematic of a system which is monostable and figure 1.5B shows a system which is bistable with two steady states. In a monostable system, irrespective of the initial conditions (values of X, Y in a two variable system), all curves representing the evolution of different variables in the system (trajectories), reach the same steady state (X_S, Y_S). In bistable systems, (figure 1.5B), depending on the initial conditions, a system can reach either of the two different steady states (X_{S1}, Y_{S1}) or (X_{S2}, Y_{S2}).

Figure 1.5 shows how in addition to deriving a steady solution for the system, knowing if a system can exhibit multistability is important to understanding a network's behaviour and its signal processing capabilities. As seen in figure 1.5, we can use time series plots to understand if a given system is monostable or multistable. The challenge with using numerical sampling of initial conditions to find all the possible steady states is that without a conclusive proof that all the steady states have been found, the results obtained will depend on the number of initial conditions that were sampled and the computational time available. Deriving an analytical solution is a more comprehensive approach to detecting the number of steady states. When this derivation is too complex or not feasible, then there are other approaches used to detect multistability in chemical reaction networks. These use interaction graphs and chemical reaction network theory (CRNT) to examine the chemical reaction structure underlying signalling networks to predict if a system has the capacity to exhibit more than one steady state.

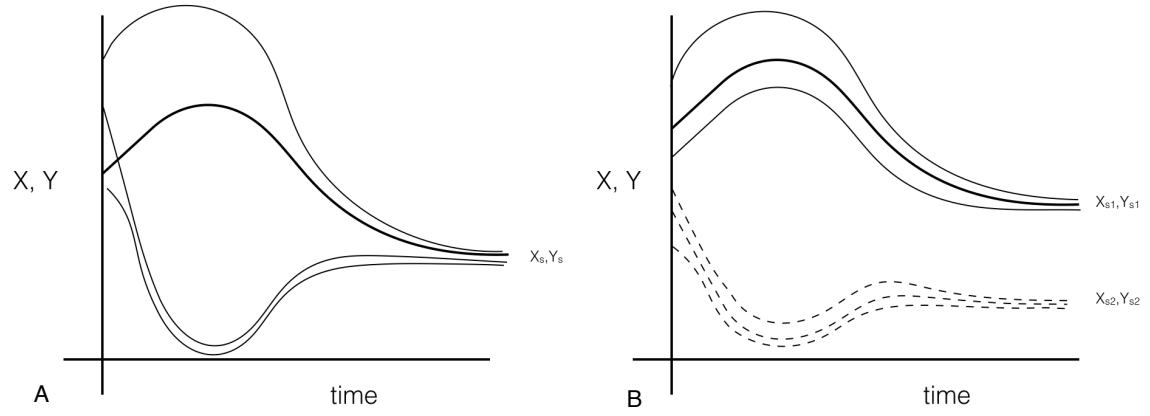


Figure 1.5: Is a representative diagram showing systems with a single steady state (X_S, Y_S) (A) and a bistable system with two stable steady states (X_{S1}, Y_{S1}) and (X_{S2}, Y_{S2}) (B). In (A), each solid line represents the system starting with different initial conditions and converging to a single steady state (X_S, Y_S). In (B), solid and dashed lines represent the system starting with different initial conditions and converging to two different steady states (X_{S1}, Y_{S1}) and (X_{S2}, Y_{S2}) respectively.

Interaction Graphs

One of the approaches used to identify if a system has multiple steady states is its interaction graph. An interaction graph is a visual representation of the reactions in a network where the nodes indicate chemical species in the reaction network and edges represent the reactions occurring between them. When building an interaction graph representing the two component signalling network described in reactions (1)-(3), we would use a positive edge to connect $\mathrm{RR_p}$ to $\mathrm{HK_p}$ if $\frac{\partial F(HK_p)}{\partial C_{RR_p}} > 0$; a negative edge if $\frac{\partial F(HK_p)}{\partial C_{RR_p}} < 0$ and no edge if $\frac{\partial F(HK_p)}{\partial C_{RR_p}} = 0$, where $\mathrm{F(HK_p)}$ is the function that represents the rate of change in concentration of $\mathrm{HK_p}$. Once we build interaction graphs for a network, Thomas [63] and Soule [64] show that multistability requires a positive circuit (see figure 1.6) where the product of signs of all edges in the circuit is positive for at least one set of species in the network. A circuit refers to a closed path along the nodes and edges of a graph where no edge is traversed more than once. If a graph does not have a positive circuit, then the system cannot be multistable. Another condition proposed by Kaufman [65] states that, multistability requires the presence of a variable nucleus with at least one edge displaying more than one sign depending on the species concentration, or a presence of two nuclei with opposite signs in a network where the sign of the nucleus with p positive circuits is given by $(-10)^{p+1}$ [66]. This nucleus in an interaction graph is defined as a union of one or more disjointed circuits which contains all vertices of the interaction graph.

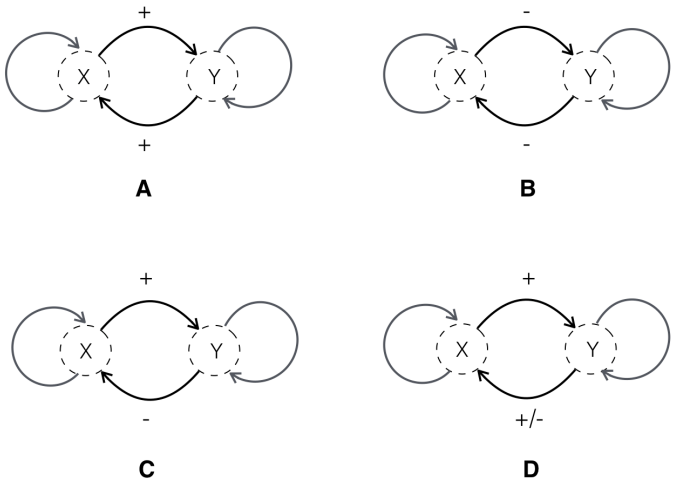


Figure 1.6: Shows circuits (representative) in an interaction graph. Positive circuits (A & B) negative circuit (C) and a circuit which can be positive or negative depending on the sign of the edge connecting Y to X (D) (Figure adapted from [62]).

Chemical Reaction Network Theory

When we model signalling networks with an assumption of mass action kinetics, the resultant equations form a chemical reaction network (CRN) [67]. We can derive a stoichiometric matrix for the reaction network and test it for injectivity [68] i.e. determine if the system can exhibit multistability [69, 70]. The presence of injectivity indicates that there are distinct positive outcomes for distinct arguments to the system. When testing these expressions for injectivity, the reactions are assumed to occur in an isothermal, homogeneous continuous flow stirred tank reactor. Another technique related to this examines a slightly distinct graph representing the CRN. It is called the species - reaction graph or SR graph [71, 72].

In this study, we used deficiency theorem implemented in CRNT (**C**hemical **R**eaction **N**etwork **T**heory) toolbox to detect multistability `https://crnt.osu.edu/LecturesOnReactionNetworks`. This theorem proposes necessary conditions for the presence, stability and multiplicity of steady states in a chemical reaction network. Chemical Reaction Network Theory Toolbox implements a parameter free approach where the assumption is that multistable CRNs have certain topological features and within each *compatibility class* there is a strict, single positive steady state with positive concentrations for all the variables in the system (`https://crnt.osu.edu/LecturesOnReactionNetworks`) [73–77] (`(http://www.jeremy-gunawardena.`

com/papers/crnt.pdf)). CRNT toolbox employs the CRN theory described in [73, 75–77](https://crnt.osu.edu/LecturesOnReactionNetworks), and was used to examine if two component and phosphorelay architectures exhibited multistability, before deriving a detailed analytical solution. For simple systems, this toolbox can also suggest example parameters where the system exhibits multistability. Following an initial test on a network's capacity for multistability, deriving a detailed analytical solution allows one to examine both necessary and sufficient conditions for multistability.

1.3 Summary

In this chapter, we introduced the background biochemistry underlying signalling in prokaryotes and the methods used to model the dynamics of prokaryotic signalling networks. Using the example of a simple two component signalling network, we discussed how ODE models are built and how numerical and analytical solutions are derived for ODE models. Building on this introduction the following chapters discuss key findings from our investigations into two component and phosphorelay architectures.

1.4 Thesis Outline

Chapter 2 examines a generic four protein phosphorelay architecture (figure 1.2A) with reverse phosphotransfer from RR $\rightarrow$ Hpt and Hpt $\rightarrow$ REC. Using ODE models representing the flow of signal molecules in this network, it discusses the dynamics of how the concentration of different phosphorylated and unphosphorylated proteins in a phosphorelay vary. By studying the temporal and steady state dynamics, it looks for specific parameter regimes where phosphorelays exhibit both hyperbolic and sigmoidal signal-response behaviour to identify the necessary and sufficient conditions for specific signal processing behaviours seen in a network. A thorough exploration of the biochemical features constituting a phosphorelay architecture is carried out by deriving 32 variants of the generic four protein phosphorelay architecture as shown in figure 1.2A. Using a combination of numerical simulations, analytical solutions and multi parameter sensitivity analysis, the role of different biochemical reactions in prescribing specific signal-response behaviours is examined. Focussing on the signal-response behaviour of networks is required for a detailed understanding of signal processing capabilities embedded in different phosphorelay architectures [16, 26, 78, 79] and is therefore explored here in detail. This in depth examination is needed to understand intracellular regulation and the evolution of

control architectures in prokaryotic cells.

Building on the investigation of different biochemical features of a phosphorelay architecture, chapter 3 examines the significance of signalling architectures employing different types of sensor kinase proteins in the network. Building models for phosphorelay architectures with multidomain histidine kinases like hybrid and unorthodox (see figure 1.2B & 1.2C), this chapter examines the dynamics that additional phosphotransfer domains on histidine kinases confer on a phosphorelay network. In comparison to sensor kinases without additional REC or Hpt domains, phosphorelays with multidomain HKs have intramolecular phosphotransfer reactions between HK→ REC and REC → Hpt domains. The chemical reaction network structure underlying phosphorelays with hybrid sensor kinases is examined using the CRNT toolbox (`https://crnt.osu.edu/CRNTWin`) [80] to test for multistability. Building on findings from this and examining systems where multiple bistable core networks share components, this chapter presents proofs that explain the correlation between the number of phosphotransfer domains on a sensor kinase, number of pathways sharing common protein components and the resultant multistability and computational logic embedded in the network.

Chapter 4 builds on theoretical findings from chapters 2 and 3 to study the distribution of different histidine kinases across genomes and annotates different HK mediated RR sequences to identify metabolic processes that different signalling architectures are involved in. This examination of proteins occurring in signalling networks is required for understanding the biological context in which different prokaryotic networks operate and potentially explain their distribution across genomes [81]. Correlating theoretical predictions from the dynamics and computational properties with the biological context in which different proteins occur could help in deriving useful insights into potential design principles underlying signalling networks in prokaryotes.

Chapter 5 summarises the key results derived from our understanding of different prokaryotic signalling architectures, their dynamics and computational properties. Building on these results, the chapter concludes by discussing how the modularity of prokaryotic signalling architectures and the availability of a number of TCS proteins could be used to build engineered control circuits for different synthetic biology applications.

2

Phosphorelays provide tunable signal processing capabilities for the cell.

Chapter Overview

Achieving a complete understanding of cellular signal transduction requires deciphering the relation between structural and biochemical features of a signalling system and the shape of the signal-response relationship it embeds. Using explicit analytical expressions and numerical simulations, this chapter examines this relation in four-layered phosphorelays which are ubiquitous in prokaryotes and some lower eukaryotes and plants. By deriving an analytical expression that relates the shape of the signal-response relationship in a relay to the kinetic rates of forward phosphorylation, reverse phosphorylation and hydrolysis reactions, we identify a set of mathematical conditions that dictate the shape of the signal-response relationship exhibited by different phosphorelay architectures. This lead to the finding that a specific phosphorelay architecture that is commonly observed in nature can satisfy these conditions in such a way that it allows plasticity (tunability) among hyperbolic and sigmoidal signal-response relationships. In particular, we find that the shape of the signal-response relationship of this relay architecture can be tuned

by altering kinetic rates and total protein concentrations in different parts of the relay. These findings are an important step towards predicting response dynamics of phosphorelays, and the nature of subsequent physiological responses that different phosphorelay architectures mediate. We are able to do this solely from their topological features and a few composite measurements like measuring the ratio of reverse and forward phosphorylation rate constants are sufficient to determine the shape of signal-response relationship that these relays exhibit. Furthermore, this study highlights the potential ways in which selective pressures on signal processing could can play a role in the evolution of observed structural and biochemical characteristics in phosphorelays.

2.1 Introduction

Signalling networks like phosphorelays allow prokaryotic cells to produce appropriate physiological responses to changes that are external and internal to the cell. Examining signal-response characteristics of these networks and how they are shaped by specific biochemical mechanisms in the network is fundamental to understanding the flow of information in a cell and to engineer cellular behaviour. Most of the studied phosphorelays to date have a sequence of four phosphotransfer reactions and contain four types of so-called two component proteins (or domains in the case of multidomain proteins) (section 1.1.2) [29, 30]. Phosphorelay networks have different lengths (number of proteins in the network), and differ with respect to the presence and location of hydrolysis and reverse phosphorylation reactions in the network. Variants of phosphorelays found in prokaryotes can contain additional RRs (> 1) [25, 82]; HKs which function as both a kinase and a phosphatase (bifunctional HKs) [25] and in some cases, could involve phosphorelays which are nested within transcriptional feedback loops [3, 36, 37]. What is the significance, if any, of these structural and biochemical features of phosphorelays? More broadly, what is the functional benefit of having a specific phosphorelay structure for the cell? To achieve a broad understanding, we built phosphorelay models with a range of alternative biochemical assumptions and parameter regimes to examine the correlation between different structural and biochemical features with their signal processing dynamics. Using both numerical simulations and analytical derivations, we evaluate the shape of signal-response relationships in all possible four-layered phosphorelay topologies (See Table 2.1) that arise when we distribute hydrolysis and reverse phosphorylation reactions on a base phosphorelay structure.

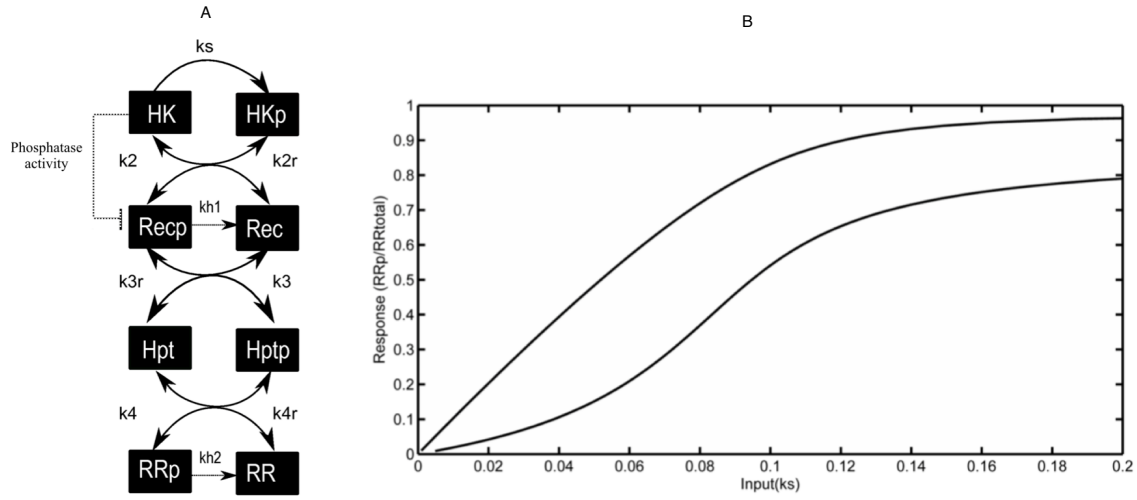


Figure 2.1: (A) Schematic representation of the general four layered phosphorelay model. Hydrolysis reactions (on aspartate residues found on REC and RR proteins only) and forward and reverse phosphorylation reactions are shown, along with the possibility of the HK being bifunctional (B) Hyperbolic and sigmoidal signal-response relationships displayed by a specific topology (topology 30 shown in Table 2.1). The two curves are obtained by choosing specific parameter sets (Appendix A). The x- and y-axis correspond to the signal and response of the system, which in the model are approximated by the HK auto-phosphorylation rate constant, k_s and by the fraction of phosphorylated RR, respectively. k_2, k_3, k_4 and k_{2r}, k_{3r}, k_{4r} are phosphotransfer and reverse phosphotransfer reactions representing the flow of phosphate groups between members of the phosphorelay (see section 2.2.1).

2.2 Modelling different phosphorelay architectures

To examine the significance of different biochemical reactions in a phosphorelay architecture, we begin with a phosphorelay which contains a monofunctional histidine kinase. We build a generic model of the four-layered phosphorelay using ordinary differential equations (ODE) and incorporate all possible combinations of reverse-phosphorylation reactions between layers (proteins) and hydrolysis reactions (i.e. encompassing all possible topologies in a four-layered relay with reverse phosphorylation and hydrolysis). In these architectures, the hydrolysis reactions are considered possible only on REC and RR as these proteins are phosphorylated on an aspartate residue (while HK and Hpt are phosphorylated on a histidine residue) and the resultant bond is known to have an inherent instability [83, 84]. Assuming a fixed phosphorelay structure containing four proteins, we derive 32 different variants (see table 2.1) of the base phosphorelay architecture shown in figure 2.1 and examine their steady state signal-response behaviour to understand the dynamical significance of different biochemical reactions occurring in the network.

2.2.1 Phosphorelays with monofunctional kinase

An ODE model for a phosphorelay contains all the four main proteins viz. histidine kinase (monofunctional) HK, a receiver protein REC, a His-containing phosphotransfer protein Hpt, and a response regulator RR. Each of these proteins can be either phosphorylated (in which case we write X_p (where X is one of the four proteins HK, REC, Hpt, and RR)) or unphosphorylated (in which case we write X). Phosphotransfer reactions occurring in this network are modelled using mass action kinetics and do not involve a complex formation. In this current model, we ignore the loss of proteins due to degradation/dilution and assume a quasi steady state concentrations of all proteins in the network. Considering all these assumptions, the final set of reactions occurring in the network are shown below:

$$\mathrm{HK} \xrightarrow{k_s} \mathrm{HK_p} \qquad \mathrm{HK_p} + \mathrm{REC} \xrightarrow{k_2} \mathrm{HK} + \mathrm{REC_p}$$

$$\mathrm{REC_p} + \mathrm{Hpt} \xrightarrow{k_3} \mathrm{REC} + \mathrm{Hpt_p} \qquad \mathrm{Hpt_p} + \mathrm{RR} \xrightarrow{k_4} \mathrm{Hpt} + \mathrm{RR_p}.$$

When we include reverse phosphotransfer reactions involving HK, REC, Hpt and RR along with hydrolysis reactions at REC_p and RR_p, phosphorelays have these additional reactions:

$$\mathrm{HK} + \mathrm{REC_p} \xrightarrow{k_{2r}} \mathrm{HK_p} + \mathrm{REC} \qquad \mathrm{REC_p} \xrightarrow{k_{h1}} \mathrm{REC}$$

$$\mathrm{REC} + \mathrm{Hpt_p} \xrightarrow{k_{3r}} \mathrm{REC_p} + \mathrm{Hpt} \qquad \mathrm{RR_p} \xrightarrow{k_{h2}} \mathrm{RR}$$

$$\mathrm{Hpt} + \mathrm{RR_p} \xrightarrow{k_{4r}} \mathrm{Hpt_p} + \mathrm{RR}.$$

From ODEs representing these reactions, we generate different variants of the four protein phosphorelay architecture by setting some of the rate constants k_{*r} or k_{h*} to zero. This helps to derive expressions for all the 32 different variants shown in table 2.1. Amongst these reactions, the minimal set of reactions which occur at rates k_2, k_3, k_4, k_s all $\neq 0$ form the base phosphorelay architecture.

	Reverse Phosphotransfer Reactions			Auto-hydrolysis Reactions	
Topology ID	$REC_p+HK \xrightarrow{k_{2r}} HK_p+REC$	$Hpt_p+REC \xrightarrow{k_{3r}} REC_p+Hpt$	$RR_p+Hpt \xrightarrow{k_{4r}} Hpt_p+RR$	$Rec_p \xrightarrow{k_{h1}} Rec + P_i$	$RR_p \xrightarrow{k_{h2}} RR + P_i$
Topology 1	1	0	0	0	1
Topology 2	0	1	0	0	1
Topology 3	0	0	1	0	1
Topology 4	1	0	1	0	1
Topology 5	1	1	0	0	1
Topology 6	0	1	1	0	1
Topology 7	0	0	0	0	1
Topology 8	1	1	1	0	1
Topology 9	1	0	0	1	0
Topology 10	0	1	0	1	0
Topology 11	0	0	1	1	0
Topology 12	1	0	1	1	0
Topology 13	1	1	0	1	0
Topology 14	0	1	1	1	0
Topology 15	0	0	0	1	0
Topology 16	1	1	1	1	0
Topology 17	1	0	0	0	0
Topology 18	0	1	0	0	0
Topology 19	0	0	1	0	0
Topology 20	1	0	1	0	0
Topology 21	1	1	0	0	0
Topology 22	0	1	1	0	0
Topology 23	0	0	0	0	0
Topology 24	1	1	1	0	0
Topology 25	1	0	0	1	1
Topology 26	0	1	0	1	1
Topology 27	0	0	1	1	1
Topology 28	1	0	1	1	1
Topology 29	1	1	0	1	1
Topology 30	0	1	1	1	1
Topology 31	0	0	0	1	1
Topology 32	1	1	1	1	1

Table 2.1: List of all possible topologies in a four-layered phosphorelay. The topologies are indicated with a binary identification code that indicates the presence (1) or absence (0) of reverse phosphotransfer reactions along the layer, and the presence (1) or absence (0) of hydrolysis reactions at layers 2 and 4 which represents proteins $\mathrm{REC_p}$ & $\mathrm{RR_p}$.

Ordinary differential equations. To simplify notations when writing differential equations for these networks, we define:

$$x_1 = [\mathrm{HK}], \qquad x_2 = [\mathrm{HK_p}], \qquad x_3 = [\mathrm{REC}], \qquad x_4 = [\mathrm{REC_p}],$$
$$x_5 = [\mathrm{Hpt}], \qquad x_6 = [\mathrm{Hpt_p}], \qquad x_7 = [\mathrm{RR}], \qquad x_8 = [\mathrm{RR_p}].$$

The resultant system of ODEs is:

$$\dot{x}_1 = -k_s x_1 - k_{2r} x_1 x_4 + k_2 x_2 x_3 \tag{26}$$

$$\dot{x}_2 = k_s x_1 + k_{2r} x_1 x_4 - k_2 x_2 x_3 \tag{27}$$

$$\dot{x}_3 = -k_2 x_2 x_3 + k_{h1} x_4 + k_{2r} x_1 x_4 + k_3 x_4 x_5 - k_{3r} x_3 x_6 \tag{28}$$

$$\dot{x}_4 = k_2 x_2 x_3 - k_{h1} x_4 - k_{2r} x_1 x_4 - k_3 x_4 x_5 + k_{3r} x_3 x_6 \tag{29}$$

$$\dot{x}_5 = -k_3 x_4 x_5 + k_{3r} x_3 x_6 + k_4 x_6 x_7 - k_{4r} x_5 x_8 \tag{30}$$

$$\dot{x}_6 = k_3 x_4 x_5 - k_{3r} x_3 x_6 - k_4 x_6 x_7 + k_{4r} x_5 x_8 \tag{31}$$

$$\dot{x}_7 = -k_4 x_6 x_7 + k_{h2} x_8 + k_{4r} x_5 x_8 \tag{32}$$

$$\dot{x}_8 = k_4 x_6 x_7 - k_{h2} x_8 - k_{4r} x_5 x_8 \tag{33}$$

Here, we see that

$$\dot{x}_1 + \dot{x}_2 = 0, \qquad \dot{x}_3 + \dot{x}_4 = 0, \qquad \dot{x}_5 + \dot{x}_6 = 0, \qquad \dot{x}_7 + \dot{x}_8 = 0.$$

As a consequence, $x_i + x_{i+1}$ is constant for $i = 1, 3, 5, 7$ and the system has four conserved quantities:

$$HK_{tot} = x_1 + x_2, \quad REC_{tot} = x_3 + x_4, \quad Hpt_{tot} = x_5 + x_6, \quad RR_{tot} = x_7 + x_8,$$

where $\mathrm{HK_{tot}}$, $\mathrm{REC_{tot}}$, $\mathrm{Hpt_{tot}}$, and $\mathrm{RR_{tot}}$ are positive constants given by the initial concentrations of the system. To ease the writing, we change the notation to

$$\overline{H} := HK_{tot}, \quad \overline{C} := REC_{tot}, \quad \overline{T} := Hpt_{tot}, \quad \overline{R} := RR_{tot}.$$

Deriving steady state solutions for the system using numerical simulations

Initially, we examined this system of ODEs by solving them numerically in Matlab [49] using its native ODE solvers, ode15s and ode23t ([50]). If we consider a plot of RR_p vs time (t), and set the duration of the simulation time to be t_{sim}. The system is considered to have reached its steady state if the concentration of RR_p has settled to a value which does not change with time i.e. $\frac{dRR_p}{dt} = 0$. We verify if $\mathrm{RR_p}$ has reached its steady state at t_{sim} by passing the values of all

variables in the ODE at time, t_{sim} as initial guesses to LMFnlsq. This function considers the Fischer's modification on the Levenberg-Marquardt algorithm to find optimal values / roots for a system of ODEs. By passing simulation data generated from these ODE solvers as initial points, we can optimise our estimate of the steady state and thus derive a more accurate approximation of the steady state signal response behaviour. Using this approach, before we had the analytical solution, we numerically derived both time-series data and steady state signal-response (k_s vs RR_p (x_8)) curves for all the 32 variants of a phosphorelay architecture.

Classifying signal-response curves using numerical simulations

To examine the nature of steady state signal-response curves and understand the significance of different reaction parameters and concentrations in a phosphorelay, we use Matlab scripts and generate signal-response curves from randomly sampled parameter values, sampled from parameter regimes defined by published experimental data (see table 1.1). We sample a total of 1000 different parameter sets and for each set, plot the steady state signal-response curve. These curves were then classified into different response types by examining the sign of the curve‘s second difference values with input (k_s) values close to (if not equal to) zero. For a sigmoidal curve, these values are positive and in the case of a hyperbolic response, they are negative. Using this approach, we derive classification data for all the 32 different architectures (see table 2.2) across all the sampled parameter sets.

			Total protein concentrations are equal		Total protein concentrations are unequal	
Topology ID	Reverse	Hydrolysis	% Hyperbolic	% Sigmoidal	% Hyperbolic	% Sigmoidal
Topology 1	1 0 0	0 1	100.000	0.000	100.000	0.000
Topology 2	0 1 0	0 1	100.000	0.000	100.000	0.000
Topology 3	0 0 1	0 1	100.000	0.000	100.000	0.000
Topology 4	1 0 1	0 1	100.000	0.000	100.000	0.000
Topology 5	1 1 0	0 1	100.000	0.000	100.000	0.000
Topology 6	0 1 1	0 1	100.000	0.000	100.000	0.000
Topology 7	0 0 0	0 1	100.000	0.000	100.000	0.000
Topology 8	1 1 1	0 1	100.000	0.000	100.000	0.000
Topology 9	1 0 0	1 0	N.R	N.R	N.R	N.R
Topology 10	0 1 0	1 0	N.R	N.R	N.R	N.R
Topology 11	0 0 1	1 0	N.R	N.R	N.R	N.R
Topology 12	1 0 1	1 0	N.R	N.R	N.R	N.R
Topology 13	1 1 0	1 0	N.R	N.R	N.R	N.R
Topology 14	0 1 1	1 0	49.267	50.733	48.133	51.867
Topology 15	0 0 0	1 0	N.R	N.R	N.R	N.R
Topology 16	1 1 1	1 0	83.200	16.800	83.267	16.733
Topology 17	1 1 1	0 0	N.R	N.R	N.R	N.R
Topology 18	1 0 0	0 0	N.R	N.R	N.R	N.R
Topology 19	0 1 0	0 0	N.R	N.R	N.R	N.R
Topology 20	0 0 1	0 0	N.R	N.R	N.R	N.R
Topology 21	1 0 1	0 0	N.R	N.R	N.R	N.R
Topology 22	1 1 0	0 0	N.R	N.R	N.R	N.R
Topology 23	0 1 1	0 0	N.R	N.R	N.R	N.R
Topology 24	0 0 0	0 0	N.R	N.R	N.R	N.R
Topology 25	1 0 0	1 1	100.000	0.000	100.000	0.000
Topology 26	0 1 0	1 1	98.467	1.533	97.867	2.133
Topology 27	0 0 1	1 1	100.000	0.000	100.000	0.000
Topology 28	1 0 1	1 1	100.000	0.000	100.000	0.000
Topology 29	1 1 0	1 1	99.667	0.333	99.867	0.133
Topology 30	0 1 1	1 1	55.400	44.600	57.133	42.867
Topology 31	0 0 0	1 1	100.000	0.000	100.000	0.000
Topology 32	1 1 1	1 1	87.800	12.200	88.800	11.200

Table 2.2: Signal-response relationship classifications for the 32 different topologies (where only 18 are responsive) using different classification and sampling schemes, i.e. equal (HK_{total}=REC_{total}=Hpt_{total}=RR_{total}) or unequal ($HK_{\text{total}} \neq REC_{\text{total}} \neq Hpt_{\text{total}} \neq RR_{\text{total}}$) total concentrations, and assuming that the phosphorelay is regulated by a monofunctional kinase. N.R. indicates that the corresponding topology is non responsive. We used numerical simulations to derive the steady state values plotted on the signal-response curve (see section 2.2.1). Classification into different response types has been done by examining the sign of the second difference values of the signal-response curve at low input (k_s) values. In these topologies, the remaining percentage of curves are categorised as linear.

From this, we find that topologies 14, 16, 30 and 32 are tunable with a capacity for exhibiting both sigmoidal and hyperbolic behaviour depending on the parameters. To investigate this further, and understand specific relationships that exist between different reaction rates and protein concentrations, we derived an explicit analytical solution for the system.

Deriving steady state solutions of the system analytically

The steady states of the system are found by setting the derivatives, $\dot{x}_i$ of the concentrations to zero, that is, $\dot{x}_i = 0$. By equating the right-hand side of the ODEs in equations (26-33) to zero we obtain a system of polynomial equations in the concentrations x_i. Due to the existence of conserved amounts, some equations are redundant. For instance, the first and second steady-state equations are

$$0 = -k_s x_1 - k_{2r} x_1 x_4 + k_2 x_2 x_3, \qquad 0 = k_s x_1 + k_{2r} x_1 x_4 - k_2 x_2 x_3.$$

One equation is minus the other, and hence, if one of them is fulfilled then so is the other. This happens because $x_1 + x_2$ is conserved. In total, four of the steady-state equations are redundant and must be replaced by the corresponding conservation equations. The steady states of the system are thus given as the solutions to the following system of equations:

$$\overline{H} = x_1 + x_2 \tag{34}$$
$$\overline{C} = x_3 + x_4 \tag{35}$$
$$\overline{T} = x_5 + x_6 \tag{36}$$
$$\overline{R} = x_7 + x_8 \tag{37}$$
$$0 = k_s x_1 + k_{2r} x_1 x_4 - k_2 x_2 x_3 \tag{38}$$
$$0 = k_2 x_2 x_3 - k_{h1} x_4 - k_{2r} x_1 x_4 - k_3 x_4 x_5 + k_{3r} x_3 x_6 \tag{39}$$
$$0 = k_3 x_4 x_5 - k_{3r} x_3 x_6 - k_4 x_6 x_7 + k_{4r} x_5 x_8 \tag{40}$$
$$0 = k_4 x_6 x_7 - k_{h2} x_8 - k_{4r} x_5 x_8. \tag{41}$$

Rearrangement of the steady-state equations.

The set of equations (34)-(41) can be replaced by another system of equations that is easier to interpret. We change equations (38)-(41) by linear combinations of them. This process does not change the set of solutions to the system. Specifically, we replace:

(40) by (40)+(41), (39) by (39)+(40)+(41), (38) by (38)+(39)+(40)+(41),

and leave (41) as it is. This results in the following equivalent system of equations:

$$\overline{H} = x_1 + x_2 \tag{42}$$
$$\overline{C} = x_3 + x_4 \tag{43}$$
$$\overline{T} = x_5 + x_6 \tag{44}$$
$$\overline{R} = x_7 + x_8 \tag{45}$$
$$0 = k_s x_1 - k_{h1} x_4 - k_{h2} x_8 \tag{46}$$
$$0 = k_2 x_2 x_3 - k_{h1} x_4 - k_{2r} x_1 x_4 - k_{h2} x_8 \tag{47}$$
$$0 = k_3 x_4 x_5 - k_{3r} x_3 x_6 - k_{h2} x_8 \tag{48}$$
$$0 = k_4 x_6 x_7 - k_{h2} x_8 - k_{4r} x_5 x_8, \tag{49}$$

where (34)-(37) are identical to (42)-(45), and (49) is identical to (41).

Using the above equations, and applying a recursive technique of substituting values of different variables, we derive expressions for the concentration (x_i) of all proteins in the network at steady state as a function of x_8. This implies that if the value of x_8 at steady state is known, then so are the values of $x_1, \ldots, x_7$:

$$x_7 = \overline{R} - x_8 \tag{50}$$
$$x_5 = \frac{k_4 \overline{T} x_7 - k_{h2} x_8}{k_4 x_7 + k_{4r} x_8} \tag{51}$$
$$x_6 = \overline{T} - x_5 \tag{52}$$
$$x_3 = \frac{k_3 x_5 \overline{C} - k_{h2} x_8}{k_3 x_5 + k_{3r} x_6} \tag{53}$$
$$x_4 = \overline{C} - x_3 \tag{54}$$
$$x_1 = \frac{k_2 x_3 \overline{H} - k_{h1} x_4 - k_{h2} x_8}{k_2 x_3 + k_{2r} x_4} \tag{55}$$
$$x_2 = \overline{H} - x_1 \tag{56}$$

The steady values derived from these expressions are only valid when they are positive. To ensure this, we derive bounds for x_8 values where all variables in the system are positive and at steady state. By iteratively substituting values of different variables in equations (50 - 56), we identify bounds for x_8, where values of x_8 are in the interval I, where $I = (0, \alpha)$ (see A.2, supplementary text S2, [85]). This ensures that all

steady state protein concentrations are positive. We can derive the value for α using equation (57). α is the first positive root of the degree-2 polynomial:

$$q_2(x_8) := a_2 x_8^2 + a_1 x_8 + a_0 \tag{57}$$

where

$$\begin{aligned}
a_2 &= k_{h2}(k_4 - k_{4r})(k_{h1} + k_2\overline{H}) + k_{h2}^2(k_3 - k_{3r}) + k_{h2}(k_3k_4 - k_{3r}k_{4r})\overline{T} \\
a_1 &= -k_{h1}(k_{3r}\overline{C}(k_{h2} + k_{4r}\overline{T}) + k_{h2}k_4\overline{R}) - k_{h2}(k_2\overline{H}(k_4\overline{R} + k_3\overline{C}) + k_3k_4\overline{T}\cdot\overline{R}) \\
&\quad -k_2k_3k_4\overline{H}\cdot\overline{C}\cdot\overline{T} \\
a_0 &= k_2k_3k_4\overline{H}\cdot\overline{C}\cdot\overline{T}\cdot\overline{R}
\end{aligned}$$

From equations (34)-(41), we also can derive an expression for k_s in terms of x_8:

$$k_s = \frac{k_{h1}x_4 + k_{h2}x_8}{x_1}.$$

If we express x_4 and x_1 in terms of x_8 using (50)-(56), we obtain that the exact analytical expression relating k_s and x_8 is:

$$k_s = f(x_8) = \frac{x_8 p_1(x_8) p_2(x_8)}{q_1(x_8) q_2(x_8)} \tag{58}$$

with $q_2(x_8)$ given as in (57) and

$$\begin{aligned}
q_1(x) &= (k_{h2}(k_3 - k_{3r}) + (k_3k_4 - k_{3r}k_{4r})\overline{T})x - k_3k_4\overline{R}\cdot\overline{T}, \\
p_1(x) &= k_{h2}\big(k_{h1}(k_4 - k_{4r}) + k_{h2}(k_3 - k_{3r}) + (k_3k_4 - k_{3r}k_{4r})\overline{T}\big)x \\
&\quad - k_4k_{h2}(k_{h1} + k_3\overline{T})\overline{R} - k_{3r}k_{h1}(k_{h2} + k_{4r}\overline{T})\overline{C}, \\
p_2(x) &= k_{h2}(k_2 - k_{2r})(k_4 - k_{4r})x^2 + k_2k_3k_4\overline{R}\cdot\overline{C}\cdot\overline{T} + \\
&\quad \big(k_{h2}(\ k_4(k_{2r} - k_2)\overline{R} + (k_{2r}k_{3r} - k_2k_3)\overline{C}\ ) + (k_{2r}k_{3r}k_{4r} - k_2k_3k_4)\overline{C}\cdot\overline{T}\big)x.
\end{aligned}$$

This function is well defined for x_8 is in $I = (0, \alpha)$, that is, it is positive and continuous. When x_8 approaches α (the upper bound of the interval I), then k_s tends to infinity (the denominator of f tends to zero). Therefore, the image of f is the interval $(0, +\infty)$. Further, the function f can be differentiably extended at zero such that $f(0) = 0$. Given a rate constant k_s, there is a unique value of x_8 for which $f(x_8) = k_s$. This value is the steady-state value of x_8 corresponding to k_s, and the other steady states as shown in (50-56).

2.2.1.1 Properties of the signal-response curve.

We let φ denote the inverse of f, that is,

$$\varphi(k_s) = x_8 \qquad \text{, if } k_s = f(x_8).$$

Using the Inverse Function Theorem, the signal-response function φ is continuous and differentiable in $[0, +\infty)$. We do not have an analytical expression for φ, only for its inverse. However, most of the information required from φ can be retrieved from f:

(i) The function φ is increasing.

(ii) α is the maximal value of the response, x_8. When the activation rate k_s tends to infinity, then x_8 approaches α.

(iii) The derivative of φ at a point $k_s = k$ equals $\varphi'(k) = 1/f'(x_8)$ for $x_8 = \varphi(k)$.

(iv) The second derivative of φ at a point $k_s = k$ equals $\varphi''(k) = -f''(x_8)/f'(x_8)^3$ for $x_8 = \varphi(k)$.

For example, the derivative of the signal-response curve φ at zero is:

$$\frac{k_3 k_4 \overline{H} \cdot \overline{R} \cdot \overline{T}}{k_4 k_{h2}(k_{h1} + k_3\overline{T})\overline{R} + k_{3r} k_{h1}(k_{h2} + k_{4r}\overline{T})\overline{C}}. \tag{59}$$

Since f is an increasing function in I, we have that $\varphi'(k) > 0$ for all $k \geq 0$ and the sign of the second derivative of φ at k is *minus* the sign of the second derivative of f at $\varphi(k)$.

Practical considerations.

(v) The signal-response curve is plotted by generating points $(f(x_8), x_8)$.

(vi) The maximal response is easily computed as the first positive root of $q_2(x)$, which is a degree-2 polynomial.

(vii) Given k_s, the steady-state value of x_8 is the first positive zero of the polynomial

$$k_s q_1(x_8) q_2(x_8) - x_8 p_1(x_8) p_2(x_8).$$

The other steady-state values are obtained from x_8 and (34-41).

2.2.1.2 Classifying signal-response curves

Once we have an expression for the steady state signal-response curve, we examine the nature of response dynamics of a phosphorelay by studying its shape. The shape carries important information about the response features of the given network [39]. A function $g(x)$ that initially increases slowly and then quickly (after crossing a threshold value) before slowing down is called sigmoidal, that is, $g'(x)$ is initially increasing then decreasing or, alternatively, $g''(x)$ is first positive and then becomes negative. Unlike this, a function $g(x)$ that increases at a slower and slower rate is called hyperbolic, that is the derivative $g'(x)$ of g is decreasing or, alternatively, the second derivative is negative, $g''(x) < 0$.

It is difficult in general to establish if a curve is sigmoidal or hyperbolic (or none of these) and here, we used a simple test to identify if φ is sigmoidal or hyperbolic, by considering the second derivative of the signal-response curve (see section 2.2.1.2). If the second derivative of φ at zero is positive, then the first derivative grows indicating that the curve will likely be sigmoidal. If, on the contrary, the second derivative of φ at zero is negative, then the curve is likely to be hyperbolic. This test is a good indicator of the shape of φ, but it is important to note that this test only considers the behaviour near zero. By building algorithms that examine sign changes in the second derivative curves on the entire signal-response curve and comparing it with a script that only examines the second derivative sign at zero, we find that the two results do not vary much. This means that the classification based on the sign of second derivative at zero is reasonable.

We compute $\varphi''(0)$ using item (iv) above and $\varphi(0) = 0$ using Maple. We find that the sign of $\varphi''(0)$ agrees with the sign of:

$$\begin{aligned} S = &- k_{h1}(k_{4r}\overline{T} + k_{h2})\Big(k_{3r}(k_2k_3k_4 - k_2k_{3r}k_{4r} + k_{2r}k_{3r}k_{4r})\overline{C}\cdot\overline{T} + k_4k_{2r}k_{3r}k_{h2}\overline{R} \\ &+ k_{h2}(k_2k_3 - k_2k_{3r} + k_{2r}k_{3r})(k_4\overline{R} + k_{3r}\overline{C})\Big)\overline{H}\cdot\overline{C} \\ &k_{2r}k_4k_{h2}(k_3k_{3r}(k_{4r}\overline{T} + k_{h2})\overline{C}\cdot\overline{T} + k_4k_{h2}(k_3\overline{T} + k_{h1})\overline{R})\overline{H}\cdot\overline{R} \\ &- (k_{3r}k_{h1}(k_{4r}\overline{T} + k_{h2})\overline{C} + k_4k_{h2}(k_3\overline{T} + k_{h1})\overline{R})^2. \end{aligned} \tag{60}$$

Identifying necessary conditions for sigmoidal signal-response curves

As discussed earlier, if $S > 0$, then the signal-response curve is classified as sigmoidal and if $S < 0$, then it is classified as hyperbolic. If the blue terms in equation (60) are all positive, then the curve is hyperbolic and hence only if some of the highlighted

blue terms are negative, is $S > 0$. Note that the negative terms in S are multiplied by k_{h1}. Therefore, *necessary* conditions for φ to be sigmoidal are

$$k_{h1} > 0 \quad \text{and} \quad (k_2k_3k_4 - k_2k_{3r}k_{4r} + k_{2r}k_{3r}k_{4r} < 0 \quad \text{or} \quad k_2k_3 - k_2k_{3r} + k_{2r}k_{3r} < 0).$$

In particular, if $k_{3r} = 0$ then sigmoidality cannot occur. By inspecting in detail the two blue terms, we see that

$$k_2k_3 - k_2k_{3r} + k_{2r}k_{3r} = k_2(k_3 - k_{3r}) + k_{2r}k_{3r} = k_2k_3 + (k_{2r} - k_2)k_{3r},$$
$$k_2k_3k_4 - k_2k_{3r}k_{4r} + k_{2r}k_{3r}k_{4r} = k_2(k_3k_4 - k_{3r}k_{4r}) + k_{2r}k_{3r}k_{4r} = k_2k_3k_4 + (k_{2r} - k_2)k_{3r}k_{4r}.$$

We conclude that necessary conditions for φ to be sigmoidal are:

$$k_{h1} > 0 \quad \text{and} \quad k_2 > k_{2r} \quad \text{and} \quad (k_{3r} > k_3 \quad \text{or} \quad k_{3r}k_{4r} > k_3k_4). \tag{61}$$

Further, we conclude the following from an analysis of the expression of S:

- If $\overline{H}$ or k_2 are very small, then S is negative and hence φ is hyperbolic.
- If $k_{h2} = 0$, then the sign of S agrees with the sign of

$$\overline{H}(k_2k_3k_4 - k_2k_{3r}k_{4r} + k_{2r}k_{3r}k_{4r}) + k_{3r}k_{4r}k_{h1}.$$

  In this case, necessary conditions for φ to be sigmoidal are:

$$k_{h1} \neq 0 \quad \text{and} \quad k_2 > k_{2r} \quad \text{and} \quad k_{3r}k_{4r} > k_3k_4.$$

Identifying necessary conditions for hyperbolic signal-response curves

It can also be shown (see Appendix B.1.2) that if:

$$k_2 - k_{2r} > 0, \qquad \text{and} \qquad (k_3 - k_{3r})k_{h2}(k_4\overline{R} + k_{3r}\overline{C}) + k_{3r}(k_3k_4 - k_{3r}k_{4r})\overline{C}\cdot\overline{T} > 0, \tag{62}$$

then the second derivative of φ strictly decreases over I and hence the curve is hyperbolic (that is, not only the second derivative at zero indicates so). These two inequalities are fulfilled if

$$k_2 > k_{2r}, \qquad k_3 > k_{3r}, \qquad k_4 > k_{4r}.$$

That is, if the phosphorelay rate constants are larger than their reverse counterparts then the curve is hyperbolic. However, the curve can be hyperbolic without these

inequalities being fulfilled.

Total protein concentrations

By expressing the term S in (60) as a polynomial which is a function of one of the total amounts, we can observe that for some parameter values, varying the total amounts can also change the system's response from sigmoidal (S positive) to hyperbolic (S negative) and vice versa. The results of the last column in the table hold true for any choice of parameters that make the leading or independent coefficients for the term S expressed in terms of total protein concentration, positive. This has been summarised in table 2.3.

	Degree	Independent coefficient	Leading coefficient	Sign of S
$\overline{H}$	1	negative	positive for some parameters	$S < 0$ for $\overline{H}$ small, $S > 0$ for $\overline{H}$ large
$\overline{C}$	2	negative	positive for some parameters	$S < 0$ for $\overline{C}$ small, $S > 0$ for $\overline{C}$ large
$\overline{T}$	3	positive for some parameters	negative	$S > 0$ for $\overline{T}$ small, $S < 0$,for $\overline{C}$ large
$\overline{R}$	4	positive for some parameters	negative	$S > 0$ for $\overline{R}$ small, $S < 0$,for $\overline{C}$ large

Table 2.3: Shows the effects of varying total protein levels in different layers. The column Degree refers to the degree of the polynomial S (60), when it is derived as a function of different total concentration of proteins in the phosphorelay.

In summary, deriving an analytical solution of the system has given us specific relationships between different parameters in a system of ODEs describing phosphorelays with a monofunctional histidine kinase. Using the necessary conditions defined in section 2.2.1.2, we carry out a sampling exercise similar to section 2.2.1, and study the distribution of different response types within a biologically relevant parameter regime.

2.2.1.3 Sampling biological parameter regimes using analytical solutions

As described in section 2.2.1, we wrote Matlab scripts that examined each topology by randomly sampling a total of 1000 different parameter sets from a uniform distribution of parameters in experimentally determined parameter regimes. For each set, we derived the signal-response relationship *f*, i.e. the steady state phosphorylated x_8 (RR_{p}) concentrations corresponding to a given k_{s} over a range of values of x_8 between 0 and $0.95 \cdot \alpha$ with increments of $\frac{\alpha}{100}$. Using this steady state signal-response curve, we examine the sign of the second derivative of the curve at zero, and arrive at this distribution of different response types across all 32 topologies (shown in table 2.4). From this table, we find that topologies 1-9, 25-29 & 31 are always hyperbolic while topologies 14, 16, 30 and 32 are tunable and can exhibit more than one response behaviour depending on specific values for parameters in the model.

2.2.2 Tunable phosphorelay architectures

As discussed in sections 2.2.1 and 2.2.1.3, we find that there are several variants of phosphorelays with a capacity for exhibiting both hyperbolic and sigmoidal response. This makes them *tunable*, giving them a capacity for exhibiting different response behaviours under different parameter regimes. Based on the necessary conditions defined for a system's response behaviour (in section 2.2.1.2), we can determine a phosphorelay's current response behaviour easily from its reaction rates and total protein concentrations. Biologically, the response behaviour can be toggled through mutations which changes reaction rates and also by regulating total protein concentrations in the relay through transcriptional feedback in the cell which controls production of new proteins in the cell. When studying such tunable architectures, a common question to ask is if one behaviour is better than the other and what it means when the network responds differently in different parameter regimes. We examine this by comparing the two architectures and testing their ability to respond to changing inputs, i.e. response time (time taken to respond to changes in the input) and measure the variance in (reliability of) a relay's output depending on the response regime.

2.2.2.1 Responsiveness in topologies 14 and 30 - a comparison between sigmoidal and hyperbolic regimes

To understand the responsiveness of tunable topologies to incoming signals, we study the differences in time taken for a relay to respond to a change in input (response time) under the two regimes (sigmoidal and hyperbolic). For each topology, we

			Total protein concentrations are equal		Total protein concentrations are unequal	
Topology ID	Reverse	Hydrolysis	% Hyperbolic	% Sigmoidal	% Hyperbolic	% Sigmoidal
Topology 1	1 0 0	0 1	100.000	0.000	100.000	0.000
Topology 2	0 1 0	0 1	100.000	0.000	100.000	0.000
Topology 3	0 0 1	0 1	100.000	0.000	100.000	0.000
Topology 4	1 0 1	0 1	100.000	0.000	100.000	0.000
Topology 5	1 1 0	0 1	100.000	0.000	100.000	0.000
Topology 6	0 1 1	0 1	100.000	0.000	100.000	0.000
Topology 7	0 0 0	0 1	100.000	0.000	100.000	0.000
Topology 8	1 1 1	0 1	100.000	0.000	100.000	0.000
Topology 9	1 0 0	1 0	N.R.	N.R.	N.R.	N.R.
Topology 10	0 1 0	1 0	N.R.	N.R.	N.R.	N.R.
Topology 11	0 0 1	1 0	N.R.	N.R.	N.R.	N.R.
Topology 12	1 0 1	1 0	N.R.	N.R.	N.R.	N.R.
Topology 13	1 1 0	1 0	N.R.	N.R.	N.R.	N.R.
Topology 14	0 1 1	1 0	49.467	50.533	48.267	51.733
Topology 15	0 0 0	1 0	N.R.	N.R.	N.R.	N.R.
Topology 16	1 1 1	1 0	84.533	15.467	85.267	14.733
Topology 17	1 0 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 18	0 1 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 19	0 0 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 20	1 0 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 21	1 1 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 22	0 1 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 23	0 0 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 24	1 1 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 25	1 0 0	1 1	100.000	0.000	100.000	0.000
Topology 26	0 1 0	1 1	98.467	1.533	97.867	2.133
Topology 27	0 0 1	1 1	100.000	0.000	100.000	0.000
Topology 28	1 0 1	1 1	100.000	0.000	100.000	0.000
Topology 29	1 1 0	1 1	99.667	0.333	99.867	0.133
Topology 30	0 1 1	1 1	55.467	44.533	57.400	42.600
Topology 31	0 0 0	1 1	100.000	0.000	100.000	0.000
Topology 32	1 1 1	1 1	89.533	10.467	90.067	9.933

Table 2.4: Signal-response relationship classifications for the 32 different topologies (where only 18 are responsive) using different classification and sampling schemes, i.e. equal (HK_{total}=REC_{total}=Hpt_{total}=RR_{total}) or unequal ($HK_{\text{total}} \neq REC_{\text{total}} \neq Hpt_{\text{total}} \neq RR_{\text{total}}$) total concentrations, and assuming that the phosphorelay is regulated by a monofunctional kinase. N.R. indicates that the corresponding topology is non responsive. We use the sign of a second derivative value at zero to derive this classification data. In these topologies, the remaining percentage of curves are categorised as linear.

picked 100 random parameter sets from both sigmoidal and hyperbolic regimes and for every parameter set, we calculated f and the maximal response α as described in (57). We then ran a time course analysis where we introduced an arbitrary step size increase (decrease) of 10% , after starting the simulation with a basal signal (i.e. k_s) level of $0.2 \cdot \alpha$, $0.5 \cdot \alpha$ and $0.8 \cdot \alpha$. The switch on (off) time was calculated as the time taken for the system to reach a new steady state after the input (k_{s}) is increased (decreased) (figures 2.2 and 2.3). We use the response times from these 100 sample parameter sets and plotted them in panels E and F (see figure 2.2). We see that the mean time taken by phosphorelays with topology 14 to respond to a step increase in input (k_{s}), also known as "switch on time"is marginally higher with a wider distribution for parameters generating a sigmoidal response than hyperbolic. In the case of topology 30, mean switch on time is instead slightly higher with a wider distribution when parameters generating hyperbolic response behaviour are chosen than sigmoidal. In addition to switch on time, we also looked at "switch off time "(see figure 2.3) and find that the topologies exhibit a behaviour similar to what was seen in figure 2.2 with marginal difference in the mean response time and distribution of response times. Both switch on and switch off times represent the time taken by a phosphorelay to respond to changing inputs, and are generally called as response time. The numerical simulations carried out here used Matlab and its native ODE solvers for stiff systems (ode15s & ode23s).

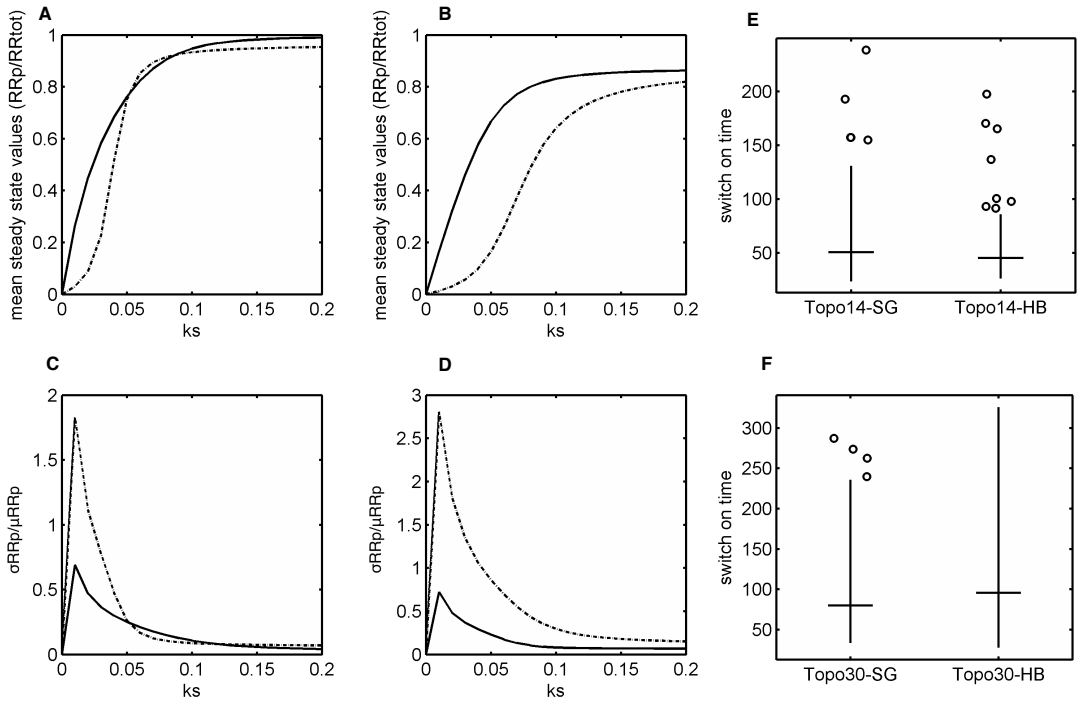


Figure 2.2: Analysis of noise and response properties of topologies 14 and 30 assuming monofunctional HK. A & B. Signal-response curve for topologies 14 (A) and 30 (B). The x-axis corresponds to the signal input to the system, which in the model is approximated by varying the HK auto-phosphorylation rate constant, k_s. The y-axis corresponds to the simulated mean of the fraction of phosphorylated RR, calculated using the PRISM model checker [86, 87]. The solid and dashed curves show the results obtained from constraining the system parameters in the hyperbolic (HB) and sigmoidal (SG) regimes respectively. C & D. Standard deviation in concentration of phosphorylated RR for topologies 14 (C) and 30 (D). The x-axis shows the signal input to the system, taken to be the HK auto-phosphorylation rate constant, k_s. The y-axis shows the standard deviation (σ_{RR_p}) over mean of the fraction of phosphorylated RR (μ_{RR_p}) at steady state, both calculated using the PRISM model checker (see section 2.2.2.2). E & F. Box plots showing the distribution of response times for topologies 14 (E) and 30 (F) as measured from hyperbolic and sigmoidal regimes. For each topology, the response time is measured for 100 randomly selected parameter sets from the hyperbolic and sigmoidal regimes.

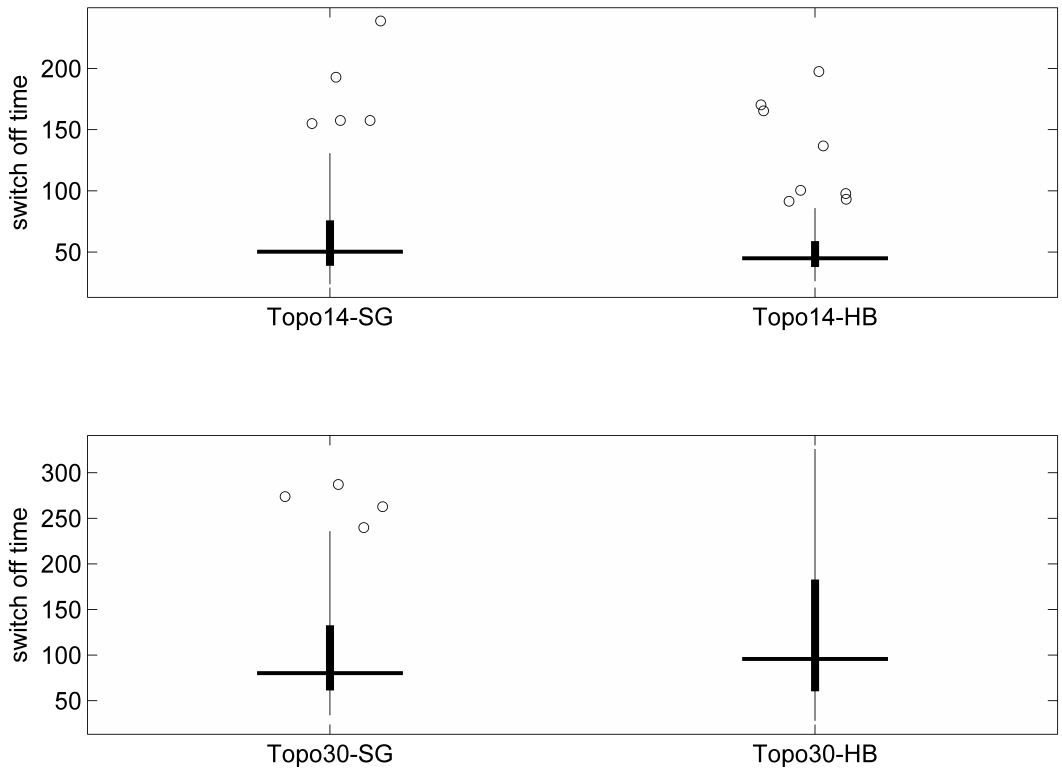


Figure 2.3: Distribution of response off times for topologies 14 (A) and 30 (B) as measured from hyperbolic (HB) and sigmoidal (SG) regimes, and shown as box plots. Response off time is defined as the time taken for the system to reach a new steady state after the input (k_s) is decreased by 10%. For each topology, the response off time is measured for 100 randomly selected parameter sets from the hyperbolic and sigmoidal regimes, based on classifications using the second derivative of the signal-response curve at zero (see section 1.2.5).

Having examined the time taken by phosphorelays in different response regimes to respond to varying inputs, we compare the reliability of response in tunable topologies when they occur in either sigmoidal or hyperbolic regimes.

2.2.2.2 Reliability of response in topologies 14 and 30 - a comparison between sigmoidal and hyperbolic regimes

When we use ODEs to model a reaction network, reactions in the model are defined in a deterministic manner. But when these networks occur in a cell, their behaviour is less deterministic due to uncertainty in their surrounding environment and internally within the network as reactions depend on the probability of collisions occurring between reactants. To understand the reliability of response in tunable topologies when they are in different response regimes, we use a probabilistic model checking [86] tool called PRISM (v4.0.3) [87] and build a Continuous Time Markov Chain (CTMC) model of the phosphorelay contain-

ing a monofunctional kinase. This probabilistic model will allow us to compare the behaviour of a tunable phosphorelay in sigmoidal and hyperbolic response regimes.

A CTMC model represents a stochastic process where the state of a system at any future time point t_{n+1} depends on its current state at t_n and is independent of the past state of the system. Each state vector of the Markov chain represents the number of molecules of all species in the system (> 0) at that point of time and evolves into the next state on the Markov chain, in accordance with a transition matrix. The state based transition matrix that we build here is defined on the basis of the stoichiometric matrix which we use to derive an ODE model (as seen in equations (26-33)) and defines a set of probabilities for transition amongst different proteins in the network. We built a CTMC model for a generic four protein phosphorelay with some of the transition probabilities equated to zero. This allows us to switch between topologies 14 and 30 (See Appendix D).

With a CTMC model which depends on the present state of the system, tools like PRISM [87] evaluate the Markov chain for all possible transition routes at any given point of time with each leading to a different state vector on the Markov chain. Using PRISM's model checking [88] tools, we can store these transition states and calculate the probability that a certain specific state occurs. The number of transition states that PRISM evaluates is directly related to the number of molecules in the system and it grows exponentially leading to a rapid explosion in the number of states. This affects the computational tractability of simulating the network. To ensure that we can simulate the PRISM model for a phosphorelay, we use a scaling factor and reduce the number of molecules in the system to 10 per protein. It has been previously shown [89] that using such a scaling ensures computational tractability and provides a solution which is qualitatively similar to the solution obtained from a more computationally intensive approach.

We converted the ODE model representing the 32 topologies into elementary reactions with probabilistic transition rates defining how the molecule numbers change after each transition (see table 2.5). All bimolecular phosphotransfer rates were transferred into probabilistic rates by dividing them with a scaling factor, $g = N_A \cdot V$, where N_A is Avogadro's number and V is a volumetric factor with dimensions M^{-1} while unimolecular reaction rates remained the same and required no scaling (s^{-1}) [90, 91]. We set V in such a way that it ensures that there are 10 molecules for each of the species in the system and the number of possible states are at a computationally tractable level.

Parameters	Mass Action Values	Dimension	Scaling by g= 53191(M^{-1})	Stochastic Values	Dimension
k_s	0 .. 0.1	s^{-1}	/1	0 .. 0.1	s^{-1}
k_2	5.32E+04	$M^{-1}s^{-1}$	/g	9.99E-01	s^{-1}
k_3	1.55E+04	$M^{-1}s^{-1}$	/g	2.91E-01	s^{-1}
k_4	2.20E+04	$M^{-1}s^{-1}$	/g	4.14E-01	s^{-1}
k_5	0.017816822	s^{-1}	/1	0.017816822	s^{-1}
k_{2r}	0	$M^{-1}s^{-1}$	/g	0.00E+00	s^{-1}
k_{3r}	6.18E+04	$M^{-1}s^{-1}$	/g	1.16E+00	s^{-1}
k_{4r}	7.49E+04	$M^{-1}s^{-1}$	/g	1.41E+00	s^{-1}
k_{2h}	0.066414053	s^{-1}	/1	6.64E-02	s^{-1}
HK_{tot}	1.85E-04	M	*g	9.86E+00	1
REC_{tot}	1.85E-04	M	*g	9.86E+00	1
Hpt_{tot}	1.85E-04	M	*g	9.86E+00	1
RR_{tot}	1.85E-04	M	*g	9.86E+00	1

Table 2.5: Converting kinetic rates used in the ODE models for topology 30 in the hyperbolic regime to the probabilistic rates used in the CTMC model described in appendix D.

Following this transformation of a kinetic ODE model into a transition matrix for a CTMC model of a phosphorelay, we simulate the model and store all the Markov state vectors for the system. We vary the phosphorylation rate that converts HK to $\mathrm{HK_p}$, i.e. k_1 in the model (see Appendix D) using the querying formalism developed in the PRISM language and calculate the mean concentration of proteins at different time points and the values of $\mathrm{RR_p}$ once the system reaches a steady state. By varying the input (k_1 in Appendix D), we derived the mean steady state RR_p values for different inputs and calculated the signal-response curves representing the system (A & B in figure 2.2) and also calculated noise (C & D in figure 2.2) in the response levels (RR_p) of topologies 14 & 30 respectively, when they were in sigmoidal and hyperbolic parameter regimes.

From figure 2.2, we see that under the sigmoidal regime, phosphorelays reach a lower maximal response level and have a high degree of standard deviation in their RR_p values at steady state for both topology 14 and 30.

In summary, we find that tunable topologies containing monofunctional kinases have more noise in their response levels when they have a sigmoidal response than when they exhibit a hyperbolic response. In addition to this, topology 30, appears to

have a shorter response time when facing varying inputs when it is in a sigmoidal response regime. With topology 14, there is no difference in the response time between the two response regimes.

So far in this chapter, we have discussed how we can derive ODE models for phosphorelays with monofunctional kinases, determine the steady state signal-response behaviours that they exhibit and the properties of tunable phosphorelay architectures in different response regimes. These models assume simplifications on phosphotransfer mechanisms, auto-dephosphorylation reactions and also assume that the total concentration of proteins are a constant. These are important simplifications which make it easier to model these networks, but differ in a few ways from the networks which function in a cell.

In the next section, we will discuss phosphorelays with a bifunctional kinase and understand how additional phosphatase activity for the kinase affects the overall dynamics in the phosphorelay.

2.2.3 Phosphorelays with bifunctional kinase

We use the same approach described in section 2.2.1 for modelling phosphorelay architectures with a bifunctional histidine kinase. A bifunctional kinase catalyses both phosphorylation and dephosphorylation of REC and REC_p [19, 21] molecules respectively and can be represented using the reaction

$$\mathrm{HK} + \mathrm{REC_p} \underset{\mathrm{k_{5r}}}{\overset{\mathrm{k_5}}{\rightleftharpoons}} \mathrm{HK} - \mathrm{REC_p} \underset{\mathrm{k_{5r}}}{\overset{\mathrm{k_6}}{\rightleftharpoons}} \mathrm{HK} + \mathrm{REC}.$$

This extends the system shown in section 2.2.1 and results in a new set of ODEs C.1.1. These ODEs differ from the original model only in the description of the rate of change for the concentrations of HK and REC, and incorporate a new species which represents the complex formed by HK and REC. Unlike phosphorelays with a monofunctional kinase, having a bifunctional kinase does not allow an explicit analytical relation. Instead of an analytical solution, we infer the existence of an analytical function relating k_s and x_8 and examine the properties of this function. Appendix C gives a detailed description of how we arrive at a signal response expression and related conditions for a phosphorelay with a bifunctional kinase. The steady state expressions for each variable x_i in the system as a function of x_8 are shown below.

$$x_7 = \overline{R} - x_8 \tag{63}$$

$$x_6 = \frac{x_8(k_{4r}\overline{T} + k_{h2})}{k_4 x_7 + k_{4r} x_8} \tag{64}$$

$$x_5 = \frac{k_4\overline{T}x_7 - k_{h2}x_8}{k_4 x_7 + k_{4r} x_8} \tag{65}$$

$$x_4 = \frac{k_{h2}x_8 + k_{3r}x_6(C - x_9)}{k_3 x_5 + k_{3r} x_6} \tag{66}$$

$$x_3 = \frac{k_3 x_5(C - x_9) - k_{h2}x_8}{k_3 x_5 + k_{3r} x_6} \tag{67}$$

$$x_2 = \frac{k_{2r}x_4(\overline{H} - x_9) + k_6 x_9 + k_{h1}x_4 + k_{h2}x_8}{k_2 x_3 + k_{2r} x_4} \tag{68}$$

$$x_1 = \frac{k_2 x_3(\overline{H} - x_9) - k_6 x_9 - k_{h1}x_4 - k_{h2}x_8}{k_2 x_3 + k_{2r} x_4} \tag{69}$$

$$x_9 = g(x_8) \tag{70}$$

x_9 represents the complex [HK-RECp]. All the other variables (x_i, where i = 1,2,3$\cdots$8) represent the same proteins as described in section 2.2.1. The expression for $g(x_8)$ can be represented as

$$G(x_8, x_9) = c_0(x_8) + c_1(x_8)x_9 + c_2(x_8)x_9^2 + c_3(x_8)x_9^3,$$

where,

$$\begin{aligned}
c_0(x_8) =& k_y x_8(k_{h2}(k_4\overline{R} + z_3 x_8) + z_1\overline{C})\Big(k_{h2}(z_3(k_{h1} + k_2\overline{H}) + z_4)x_8^2 \\
&+ (\overline{C}(k_{h1}z_1 + z_2\overline{H}) + k_{h2}k_4\overline{R}(k_2\overline{H} + k_3\overline{T} + k_{h1}))x_8 - k_2k_3k_4\overline{CHTR}\Big) \\
c_1(x_8) =& \Big((k_{2r}z_1 - z_2)\overline{C}x_8 - k_{h2}(k_2 - k_{2r})(k_4\overline{R} + z_3x_8)x_8 + k_2k_3k_4\overline{CTR}\Big)(k_3k_4\overline{TR} + z_4x_8) \\
&+ \Big((k_2k_3k_4\overline{TR} - x_8z_2)(k_{h2}(x_8z_3 + k_4\overline{R})(\overline{H} + \overline{C}) + z_1(2\overline{H} + \overline{C})\overline{C}) \\
&- k_{h2}k_2z_1(x_8z_3 + k_4\overline{R})(\overline{H} + \overline{C})x_8 - k_2k_{h2}^2x_8(x_8^2z_3^2 + k_4^2\overline{R}^2) \\
&- 2k_{h2}z_3(k_2k_4k_{h2}\overline{R} + z_1k_{h1})x_8^2 + k_{h2}x_8(k_6z_3 - z_1)(k_3k_4\overline{TR} + z_4x_8) \\
&+ (k_6z_4x_8 - 2z_1k_{h1}x_8 + k_3k_4k_6\overline{TR})(z_1\overline{C} + k_{h2}k_4\overline{R})\Big)x_8k_y \\
c_2(x_8) =& \Big(k_{h2}z_3(z_2 + k_2z_1)x_8^2 + (k_{h1}z_1^2 + (2z_2\overline{C} + k_2k_{h2}k_4\overline{R} - k_6z_4 + z_2\overline{H})z_1 \\
&- k_4k_{h2}\overline{R}(k_2k_3z_3\overline{T} - z_2))x_8 - k_3k_4\overline{RT}(k_2k_4k_{h2}\overline{R} + z_1(k_6 + 2\overline{C}k_2 + k_2\overline{H}))\Big)x_8k_y \\
&- (k_3k_4\overline{TR} + z_4x_8)(k_2k_3k_4\overline{TR} + (k_{2r}z_1 - z_2)x_8) \\
c_3(x_8) =& z_1(k_2k_3k_4\overline{TR} - z_2x_8)k_yx_8
\end{aligned}$$

with

$$z_1 := k_{3r}(k_{4r}T+k_{h2}), \quad z_2 := k_2k_3(k_4T+k_{h2}), \quad z_3 := k_{4r}-k_4, \quad z_4 := z_1-k_3(k_4T+k_{h2}),$$

2.2.4 Comparing monofunctional and bifunctional HK mediated phosphorelays

When examining the signal response expression in section C.1.2, we see that the maximal response of phosphorylated response regulator $x_8 = [\mathrm{RR_p}]$ is independent of whether the kinase is bifunctional or not. In particular, we see that its value is independent of the rate constants k_5, $k_{5\mathrm{r}}$, k_6. However, the signal-response curve in the bifunctional case is always below the signal-response curve in the monofunctional case. The expression for when the kinase is monofunctional (seen in (34)-(41)) can easily be obtained from those in (205) by setting $x_9 = 0$. It follows that if the common reactions have the same rate constants in the two cases, then for every fixed x_8, the value of x_4 (resp. x_1) in the bifunctional case is smaller (resp. larger) than in the monofunctional case. Consequently, for any rate constants k_5, $k_{5\mathrm{r}}$, k_6, we have $f_b(x_8) < f(x_8)$. This means that the signal k_s required to achieve a certain response x_8 is smaller in the monofunctional case than in the bifunctional case. Nevertheless, as k_s increases, the steady-state value of x_8 tends to the maximal response (which is the same value in both cases). Therefore, the difference between the steady-state value of x_8 in the two cases becomes negligible for large values of k_s.

This is due to sequestration of substrate in x_9. Therefore, the signal-response curve in the monofunctional case is always above the corresponding curve in the bifunctional case (for any choice of additional rate parameters). Since the maximal response is independent of the role of the kinase, a smaller signal is required to get close to the maximal response, when the kinase is monofunctional. Furthermore, if the reciprocal of the Michaelis-Menten constant 72 $k_y = k_5/(k_{5r} + k_6)$ (equation (204)) increases and k_6 is fixed, then $f_b(x_8)$ increases for a fixed x_8 (see Appendix C.2.3).

Unlike phosphotransfer seen in two component signalling networks, enzyme mediated reactions involving intermediates (complex formation) are usually modelled under the assumptions of Michaelis-Menten kinetics [92, 93]. Consider a biological catalyst (enzyme, E) binding to a substrate (S), and forming a complex (ES) (2.2.4). ES dissociates to form a product P at a rate k_p in addition to releasing the original substrate (S) and enzyme (E) back into the reactant pool at a rate k_r,

$$\mathrm{E} + \mathrm{S} \underset{\mathrm{k_r}}{\overset{\mathrm{k_f}}{\rightleftharpoons}} \mathrm{ES} \overset{\mathrm{k_p}}{\rightarrow} \mathrm{E} + \mathrm{P}$$

The rate of product formation ($\frac{dP}{dt}$) for this system can be derived as a function of substrate concentration (S) and reaction rate k_p. Assuming that the complex, ES reaches equilibrium very quickly and its concentration is always at steady state (irrespective of the concentration of substrate or product [94] molecules), we can define the rate of product formation ($\frac{dP}{dt}$) as

$$\frac{dP}{dt} = \frac{V_{max}.S}{K_m + S} \tag{71}$$

V_{max} is the maximal rate of product formation where $V_{max} = k_p.E_t$ and $E_t = E + ES$ (total enzyme concentration, assumed to be constant). K_m is known as the Michaelis constant where

$$K_m = \frac{k_r + k_p}{k_f} \tag{72}$$

It represents the strength of association between E and S in the ES complex and affects both the rate of product formation and the availability of free substrate molecules (S) in the reactant system.

In addition to this, as described in sections 2.2.1 and 2.2.1.3, we analysed the distribution of different response dynamics in phosphorelays with bifunctional kinase. Similar to our study in sections 2.2.1 and 2.2.1.3, we sampled 1000 different parameter sets within experimentally determined parameter regimes, using both numerical simulations and solving the system of ODEs analytically (classification data shown in tables 2.6 & 2.7 respectively).

			Total protein concentrations are equal		Total protein concentrations are unequal	
Topology ID	Reverse	Hydrolysis	% Hyperbolic	% Sigmoidal	% Hyperbolic	% Sigmoidal
Topology 1	1 0 0	0 1	100.00	0.00	100.00	0.00
Topology 2	0 1 0	0 1	98.20	1.80	99.40	0.60
Topology 3	0 0 1	0 1	99.90	0.10	99.90	0.10
Topology 4	1 0 1	0 1	99.90	0.10	99.80	0.20
Topology 5	1 1 0	0 1	98.70	1.30	98.80	1.20
Topology 6	0 1 1	0 1	49.60	50.40	51.20	48.80
Topology 7	0 0 0	0 1	100.00	0.00	99.90	0.10
Topology 8	1 1 1	0 1	51.10	48.90	57.60	42.40
Topology 9	1 0 0	1 0	N.R	N.R	N.R	N.R
Topology 10	0 1 0	1 0	N.R	N.R	N.R	N.R
Topology 11	0 0 1	1 0	N.R	N.R	N.R	N.R
Topology 12	1 0 1	1 0	N.R	N.R	N.R	N.R
Topology 13	1 1 0	1 0	N.R	N.R	N.R	N.R
Topology 14	0 1 1	1 0	47.20	52.80	50.90	49.10
Topology 15	0 0 0	1 0	N.R	N.R	N.R	N.R
Topology 16	1 1 1	1 0	50.80	49.20	53.20	46.80
Topology 17	1 1 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 18	1 0 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 19	0 1 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 20	0 0 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 21	1 0 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 22	1 1 0	0 0	37.900	61.900	35.700	64.100
Topology 23	0 1 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 24	0 0 0	0 0	37.8	62.2	42.6	57.1
Topology 25	1 0 0	1 1	100.00	0.00	99.80	0.20
Topology 26	0 1 0	1 1	99.60	0.40	99.00	1.00
Topology 27	0 0 1	1 1	99.90	0.10	99.50	0.50
Topology 28	1 0 1	1 1	99.90	0.10	99.60	0.40
Topology 29	1 1 0	1 1	99.50	0.50	99.00	1.00
Topology 30	0 1 1	1 1	51.60	48.40	58.30	41.70
Topology 31	0 0 0	1 1	100.00	0.00	100.00	0.00
Topology 32	1 1 1	1 1	60.80	39.20	62.00	38.00

Table 2.6: Signal-response relationship classifications for the 32 different topologies (where only 20 are responsive) using different sampling schemes, i.e. equal (HK_{total}=REC_{total}=Hpt_{total}=RR_{total}) or unequal ($HK_{\text{total}} \neq REC_{\text{total}} \neq Hpt_{\text{total}} \neq RR_{\text{total}}$) total concentrations, and assuming that the phosphorelay is regulated by a bifunctional kinase. We used numerical simulations to derive the steady state values plotted on the signal-response curve and classification into different response types was done by examining the sign of second difference values of the signal-response curve at low input (k_s) levels (see section 2.2.1). N.R. indicates that the corresponding topology is non responsive. In these topologies, the remaining percentage of curves are categorised as linear.

These tables shows that having bifunctional histidine kinase increases the number of responsive topologies with non constant response curves. For example, topologies 22 & 24 in table 2.6 and table 2.7 are responsive when the kinase has an additional phosphatase activity (unlike monofunctional HK mediated phosphorelay topologies shown in tables 2.2 and 2.4). We also see that unlike in the case of a phosphorelay containing monofunctional kinase, there are more tunable architectures here, i.e. topologies 6, 8, 22 & 24 (refer to table 2.1). In topologies 16 & 32, there is a significant drop in the percentage of hyperbolic curves when there is a bifunctional histidine kinase, along with an increase in the number of sigmoidal curves.

In the models that we discussed so far, we ignore complex formation and degradation of proteins. We also assume the absence of hydrolysis reactions at the histidine containing residues. In the next part of this thesis, we will examine how different assumptions on the mechanisms underlying this signalling network can affect our understanding of phosphorelays and more specifically, compare the criterion for both necessary and sufficient conditions for different signal-response behaviours.

2.3 Phosphorelays with different modelling assumptions

2.3.1 Model with intermediates

We extend the model given in section 2.2.1 to incorporate the formation of intermediate complexes at the phosphotransfer reactions. Intermediates refer to the complexes formed when reactants bind to each other during reactions, before the products are formed. In the reactions below, $\mathrm{Y_i}$, where $i = 1, 2, 3$ represent the intermediates formed in the system. That is, the model extended with intermediates consist of the reactions

$$\begin{aligned}
&\mathrm{HK} \xrightarrow{k_s} \mathrm{HK_p} && \mathrm{HK_p} + \mathrm{REC} \underset{k_{2ar}}{\overset{k_{2a}}{\rightleftarrows}} \mathrm{Y_1} \underset{k_{2br}}{\overset{k_{2b}}{\rightleftarrows}} \mathrm{HK} + \mathrm{REC_p} \\
&\mathrm{REC_p} \xrightarrow{k_{h1}} \mathrm{REC} && \mathrm{REC_p} + \mathrm{Hpt} \underset{k_{3ar}}{\overset{k_{3a}}{\rightleftarrows}} \mathrm{Y_2} \underset{k_{3br}}{\overset{k_{3b}}{\rightleftarrows}} \mathrm{REC} + \mathrm{Hpt_p} \\
&\mathrm{RR_p} \xrightarrow{k_{h2}} \mathrm{RR} && \mathrm{Hpt_p} + \mathrm{RR} \underset{k_{4ar}}{\overset{k_{4a}}{\rightleftarrows}} \mathrm{Y_3} \underset{k_{4br}}{\overset{k_{4b}}{\rightleftarrows}} \mathrm{Hpt} + \mathrm{RR_p}.
\end{aligned}$$

By setting some of the rate constants k_{*r} or k_{h*} to zero, we obtain different topologies involving different combinations of reverse phosphorelay and hydrolysis reactions as before.

			Total protein concentrations are equal		Total protein concentrations are unequal	
Topology ID	Reverse	Hydrolysis	% Hyperbolic	% Sigmoidal	% Hyperbolic	% Sigmoidal
Topology 1	1 0 0	0 1	100.00	0.00	100.00	0.00
Topology 2	0 1 0	0 1	98.30	1.70	98.70	1.30
Topology 3	0 0 1	0 1	100.00	0.00	100.00	0.00
Topology 4	1 0 1	0 1	100.00	0.00	100.00	0.00
Topology 5	1 1 0	0 1	98.70	1.30	98.70	1.30
Topology 6	0 1 1	0 1	50.50	47.40	47.80	52.20
Topology 7	0 0 0	0 1	100.00	0.00	100.00	0.00
Topology 8	1 1 1	0 1	52.70	45.10	51.90	48.10
Topology 9	1 0 0	1 0	N.R.	N.R.	N.R.	N.R.
Topology 10	0 1 0	1 0	N.R.	N.R.	N.R.	N.R.
Topology 11	0 0 1	1 0	N.R.	N.R.	N.R.	N.R.
Topology 12	1 0 1	1 0	N.R.	N.R.	N.R.	N.R.
Topology 13	1 1 0	1 0	N.R.	N.R.	N.R.	N.R.
Topology 14	0 1 1	1 0	48.30	50.20	43.50	55.00
Topology 15	0 0 0	1 0	N.R.	N.R.	N.R.	N.R.
Topology 16	1 1 1	1 0	52.80	45.80	43.90	51.60
Topology 17	1 0 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 18	0 1 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 19	0 0 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 20	1 0 1	0 0	N.R.	N.R.	N.R.	N.R.
Topology 21	1 1 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 22	0 1 1	0 0	43.400	34.200	38.800	34.200
Topology 23	0 0 0	0 0	N.R.	N.R.	N.R.	N.R.
Topology 24	1 1 1	0 0	40.8	31.5	46.6	26.2
Topology 25	1 0 0	1 1	100.00	0.00	100.00	0.00
Topology 26	0 1 0	1 1	99.60	0.40	98.70	1.30
Topology 27	0 0 1	1 1	99.80	0.00	100.00	0.00
Topology 28	1 0 1	1 1	99.90	0.00	100.00	0.00
Topology 29	1 1 0	1 1	99.50	0.50	98.90	1.10
Topology 30	0 1 1	1 1	52.70	46.40	52.60	47.40
Topology 31	0 0 0	1 1	100.00	0.00	100.00	0.00
Topology 32	1 1 1	1 1	62.10	37.80	55.80	44.20

Table 2.7: The results of signal-response relationship classification for the 32 different topologies (where only 20 are responsive) using different sampling schemes (equal or different total protein concentrations at different layers) equal (HK_{total}=REC_{total}=Hpt_{total}=RR_{total}) or unequal ($HK_{\text{total}} \neq REC_{\text{total}} \neq Hpt_{\text{total}} \neq RR_{\text{total}}$) total concentrations, and assuming that the phosphorelay is regulated by a bifunctional kinase. N.R. indicates that the corresponding topology is non responsive. We derived steady state analytical expressions for the steady state values plotted on the signal-response curve. The classification of these curves into different response types was done by examining the sign of the expression representing the second derivative of k_s as a function of RR_p, when $RR_p = 0$. In these topologies, the remaining percentage of curves are categorised as linear.

Ordinary differential equations.

We model protein concentrations in the system using ordinary differential equations (ODEs). To simplify the notation, we define:

$$\begin{aligned} &x_1 = [\mathrm{HK}], && x_2 = [\mathrm{HK_p}], && x_3 = [\mathrm{REC}], && x_4 = [\mathrm{REC_p}], \\ &x_5 = [\mathrm{Hpt}], && x_6 = [\mathrm{Hpt_p}], && x_7 = [\mathrm{RR}], && x_8 = [\mathrm{RR_p}], \\ &x_9 = [Y_1], && x_{10} = [Y_2], && x_{11} = [Y_3]. \end{aligned}$$

The dynamics of the concentrations in time is modelled with a system of ODEs:

$$\dot{x}_1 = -k_s x_1 - k_{2br} x_1 x_4 + k_{2b} x_9 \tag{73}$$

$$\dot{x}_2 = k_s x_1 + k_{2ar} x_9 - k_{2a} x_2 x_3 \tag{74}$$

$$\dot{x}_3 = -k_{2a} x_2 x_3 + k_{h1} x_4 + k_{2ar} x_9 + k_{3b} x_{10} - k_{3br} x_3 x_6 \tag{75}$$

$$\dot{x}_4 = k_{2b} x_9 - k_{h1} x_4 - k_{2br} x_1 x_4 - k_{3a} x_4 x_5 + k_{3ar} x_{10} \tag{76}$$

$$\dot{x}_5 = -k_{3a} x_4 x_5 + k_{3ar} x_{10} + k_{4b} x_{11} - k_{4br} x_5 x_8 \tag{77}$$

$$\dot{x}_6 = k_{3b} x_{10} - k_{3br} x_3 x_6 - k_{4a} x_6 x_7 + k_{4ar} x_{11} \tag{78}$$

$$\dot{x}_7 = -k_{4a} x_6 x_7 + k_{h2} x_8 + k_{4ar} x_{11} \tag{79}$$

$$\dot{x}_8 = k_{4b} x_{11} - k_{h2} x_8 - k_{4br} x_5 x_8 \tag{80}$$

$$\dot{x}_9 = k_{2a} x_2 x_3 - k_{2ar} x_9 + k_{2br} x_1 x_4 - k_{2b} x_9 \tag{81}$$

$$\dot{x}_{10} = k_{3a} x_4 x_5 - k_{3ar} x_{10} - k_{3b} x_{10} + k_{3br} x_3 x_6 \tag{82}$$

$$\dot{x}_{11} = k_{4a} x_6 x_7 - k_{4ar} x_{11} + k_{4br} x_5 x_8 - k_{4b} x_{11}. \tag{83}$$

The system has four conserved quantities:

$$\begin{aligned} &\mathrm{HK_{tot}} = x_1 + x_2 + x_9, \quad \mathrm{REC_{tot}} = x_3 + x_4 + x_9 + x_{10}, \\ &\mathrm{Hpt_{tot}} = x_5 + x_6 + x_{10} + x_{11}, \quad \mathrm{RR_{tot}} = x_7 + x_8 + x_{11}, \end{aligned}$$

where HK_{tot}, REC_{tot}, Hpt_{tot}, and RR_{tot} are positive constants given by the initial concentrations of the system. As before, we write

$$\overline{H} := HK_{tot}, \quad \overline{C} := REC_{tot}, \quad \overline{T} := Hpt_{tot}, \quad \overline{R} := RR_{tot}.$$

Steady-state equations.
We proceed as in the monofunctional case and conclude that the steady states of the system are given as the solutions to the following system of equations:

$$\overline{H} = x_1 + x_2 + x_9 \tag{84}$$

$$\overline{C} = x_3 + x_4 + x_9 + x_{10} \tag{85}$$

$$\overline{T} = x_5 + x_6 + x_{10} + x_{11} \tag{86}$$

$$\overline{R} = x_7 + x_8 + x_{11} \tag{87}$$

$$0 = -k_s x_1 - k_{2br} x_1 x_4 + k_{2b} x_9 \tag{88}$$

$$0 = -k_{2a} x_2 x_3 + k_{h1} x_4 + k_{2ar} x_9 + k_{3b} x_{10} - k_{3br} x_3 x_6 \tag{89}$$

$$0 = k_{2b} x_9 - k_{h1} x_4 - k_{2br} x_1 x_4 - k_{3a} x_4 x_5 + k_{3ar} x_{10} \tag{90}$$

$$0 = -k_{3a} x_4 x_5 + k_{3ar} x_{10} + k_{4b} x_{11} - k_{4br} x_5 x_8 \tag{91}$$

$$0 = k_{3b} x_{10} - k_{3br} x_3 x_6 - k_{4a} x_6 x_7 + k_{4ar} x_{11} \tag{92}$$

$$0 = -k_{4a} x_6 x_7 + k_{h2} x_8 + k_{4ar} x_{11} \tag{93}$$

$$0 = k_{4b} x_{11} - k_{h2} x_8 - k_{4br} x_5 x_8 \tag{94}$$

Hyperbolic vs. sigmoidal.
It is not so straightforward in this case to obtain an analytical description of the signal-response curve. Therefore, we adopt a direct route to the computation of the sign of the second derivative at zero.
When $k_s = 0$, then the steady state of the system equals

$$(x_1, \ldots, x_{11}) = (\overline{H}, 0, \overline{C}, 0, \overline{T}, 0, \overline{R}, 0, 0, 0, 0).$$

We want to find the derivative of the response x_8 with respect to k_s at $k_s = 0$ at steady state. To this end, we do the following steps:

1. We take the derivative with respect to k_s of both sides of the steady-state equations. We obtain new equations, where $p_i = \frac{\partial x_i}{\partial k_s}$:

$$0 = -x_1 - k_s p_1 - k_{2br} p_1 x_4 - k_{2br} x_1 p_4 + k_{2b} p_9, \tag{95}$$

$$0 = -k_{2a} p_2 x_3 - k_{2a} x_2 p_3 + k_{2ar} p_9 - k_{3br} p_3 x_6 - k_{3br} x_3 p_6 + k_{3b} p_{10} + k_{h1} p_4, \tag{96}$$

$$0 = -k_{2br} p_1 x_4 - k_{2br} x_1 p_4 + k_{2b} p_9 - k_{3a} p_4 x_5 - k_{3a} x_4 p_5 + k_{3ar} p_{10} - k_{h1} p_4, \tag{97}$$

$$0 = -k_{3a}p_4x_5 - k_{3a}x_4p_5 + k_{3ar}p_{10} - k_{4br}p_5x_8 - k_{4br}x_5p_8 + k_{4b}p_{11}, \tag{98}$$

$$0 = -k_{3br}p_3x_6 - k_{3br}x_3p_6 + k_{3b}p_{10} - k_{4a}p_6x_7 - k_{4a}x_6p_7 + k_{4ar}p_{11} \tag{99}$$

$$0 = -k_{4a}p_6x_7 - k_{4a}x_6p_7 + k_{4ar}p_{11} + k_{h2}p_8, \tag{100}$$

$$0 = -k_{4br}p_5x_8 - k_{4br}x_5p_8 + k_{4b}p_{11} - k_{h2}p_8, \tag{101}$$

$$0 = p_1 + p_2 + p_9 \tag{102}$$

$$0 = p_3 + p_4 + p_9 + p_{10}, \tag{103}$$

$$0 = p_5 + p_6 + p_{10} + p_{11}, \tag{104}$$

$$0 = p_7 + p_8 + p_{11}. \tag{105}$$

2. We substitute, in the equations above, the steady-state value when $k_s = 0$ and obtain:

$$0 = -\overline{H} - k_s p_1(0) - k_{2br}p_1(0) - k_{2br}\overline{H}p_4(0) + k_{2b}p_9(0),$$
$$0 = -k_{2a}\overline{C}p_2(0) - k_{2a}p_3(0) + k_{2ar}p_9(0) - k_{3br}p_3(0) - k_{3br}\overline{C}p_6(0) + k_{3b}p_{10}(0) + k_{h1}p_4(0),$$
$$0 = -k_{2br}p_1(0) - k_{2br}\overline{H}p_4(0) + k_{2b}p_9(0) - k_{3a}\overline{T}p_4(0) - k_{3a}p_5(0) + k_{3ar}p_{10}(0) - k_{h1}p_4(0),$$
$$0 = -k_{3a}\overline{T}p_4(0) - k_{3a}p_5(0) + k_{3ar}p_{10}(0) - k_{4br}p_5(0) - k_{4br}\overline{T}p_8(0) + k_{4b}p_{11}(0),$$
$$0 = -k_{3br}p_3(0) - k_{3br}\overline{C}p_6(0) + k_{3b}p_{10}(0) - k_{4a}\overline{R}p_6(0) - k_{4a}p_7(0) + k_{4ar}p_{11}(0),$$
$$0 = -k_{4a}\overline{R}p_6(0) - k_{4a}p_7(0) + k_{4ar}p_{11}(0) + k_{h2}p_8(0),$$
$$0 = -k_{4br}p_5(0) - k_{4br}\overline{T}p_8(0) + k_{4b}p_{11}(0) - k_{h2}p_8(0),$$
$$0 = p_1(0) + p_2(0) + p_9(0),$$
$$0 = p_3(0) + p_4(0) + p_9(0) + p_{10}(0),$$
$$0 = p_5(0) + p_6(0) + p_{10}(0) + p_{11}(0),$$
$$0 = p_7(0) + p_8(0) + p_{11}(0).$$

This system is linear in $p_1(0), \ldots, p_{11}(0)$ and hence the derivatives of x_i at $k_s = 0$ can be found by solving the system. We have solved it using Maple. In particular, we have obtained that

$$p_8(0) = \frac{k_{4a}k_{3b}k_{3a}k_{4b}\overline{H} \cdot \overline{R} \cdot \overline{T}}{denom} \tag{106}$$

Where

$$\begin{aligned} denom &= k_{h2}k_{4a}k_{3b}k_{4b}\overline{R}(k_{h1}+k_{3a}\overline{T}) \\ &\quad +k_{h1}k_{3ar}(k_{h2}k_{4ar}(k_{3br}\overline{C}+k_{4a}\overline{R})+k_{3br}\overline{C}(k_{h2}k_{4b}+k_{4br}k_{4ar}\overline{T})) \end{aligned}$$

This is the derivative of the signal-response curve at $k_s = 0$.

3. We repeat the steps above one more time: we compute the derivative with respect to k_s of the above equations (step 1). We evaluate the resulting equations at the steady state for $k_s = 0$ and at $p_i = p_i(0)$. We obtain a linear system in the second derivatives of x_i at $k_s = 0$ which can be solved in Maple. As a result, we obtain the second derivative of x_8 with respect to k_s at $k_s = 0$ as desired.

Before showing what the second derivative of x_8 with respect to k_s at $k_s = 0$ is, it is convenient to introduce new parameters. For $i = 2, 3, 4$, let

$$k_{iy} = \frac{k_{ia}}{k_{iar}+k_{ib}}, \qquad k_{iyr} = \frac{k_{ibr}}{k_{iar}+k_{ib}}, \qquad k_i = k_{ib}k_{iy}, \qquad k_{ir} = k_{iar}k_{iyr}.$$

For an interpretation of these constants see below. In particular, we take k_i, k_{ir} to be the rates of forward and reverse phosphorylation at each layer. With these new constants, we obtain that the derivative of x_8 with respect to k_s at $k_s = 0$, that is (106), becomes

$$p_8(0) = \frac{k_3k_4\overline{H}\cdot\overline{R}\cdot\overline{T}}{k_4k_{h2}(k_{h1}+k_3\overline{T})\overline{R}+k_{3r}k_{h1}(k_{h2}+k_{4r}\overline{T})\overline{C}}. \tag{107}$$

This expression is identical to the first derivative of x_8 with respect to k_s at $k_s = 0$ for the model without intermediates, as given in (59). Similarly, the sign of the second derivative of x_8 with respect to k_s at $k_s = 0$ equals the sign of:

$$\begin{aligned} S_y = {} & S - CHk_2k_3k_{h1}(k_{4y}(k_{4r}T+k_{h2})+Tk_4k_{4yr})(Rz_2+CTz_1) \\ & - CHk_2k_{h1}(z_2-Tz_1)((Cz_1+Rk_{h2}k_4)k_{3y}+C(k_{4r}T+k_{h2})k_3k_{3yr}) \\ & - C(k_{h1}z_1(C-H)+z_3)(H(Cz_1+Rk_{h2}k_4)(k_2k_{2yr}+k_{2r}k_{2y})+(Ck_{h1}z_1+z_3)k_{2y}) \end{aligned} \tag{108}$$

where S is given in (60) and

$$z_1 := k_{3r}(k_{4r}T+k_{h2}), \quad z_2 := k_{h2}(k_4R+k_{3r}C), \quad z_3 := Rk_{h2}k_4(k_3T+k_{h1}).$$

Recall that the condition for sigmoidality is $S_y > 0$. We have marked in blue the only terms that can cause the term S_y to be positive. Namely, if S is negative (that is, the model without intermediates is hyperbolic), $C > H$ and $z_2 > Tz_1$, then the model with intermediates is hyperbolic as well.
We deduce easily that

- If $k_{h1} = 0$, then sigmoidality cannot occur.
- If $k_{3r} = 0$ then $z_1 = 0$ and $S < 0$, and hence sigmoidality cannot occur.

Interpretation of the new rate constants.

The rate constants k_{iy}, k_{iyr} are the reciprocal of the Michaelis-Menten constants of each intermediate $\mathrm{Y_i}$ in each direction. These are the coefficients of the expression in $x_1, \ldots, x_8$ obtained by imposing $\dot{x}_9 = \dot{x}_{10} = \dot{x}_{11} = 0$ and solving for x_9, x_{10}, x_{11}. In particular, at steady state we have:

$$\begin{aligned}
x_9 &= \frac{k_{2a}}{k_{2ar} + k_{2b}} x_2 x_3 + \frac{k_{2br}}{k_{2ar} + k_{2b}} x_1 x_4 = k_{2y} x_2 x_3 + k_{2yr} x_1 x_4 \\
x_{10} &= \frac{k_{3a}}{k_{3ar} + k_{3b}} x_4 x_5 + \frac{k_{3br}}{k_{3ar} + k_{3b}} x_3 x_6 = k_{3y} x_4 x_5 + k_{3yr} x_3 x_6 \\
x_{11} &= \frac{k_{4a}}{k_{4ar} + k_{4b}} x_6 x_7 + \frac{k_{4br}}{k_{4ar} + k_{4b}} x_5 x_8 = k_{4y} x_6 x_7 + k_{4yr} x_5 x_8.
\end{aligned}$$

If we plug these values into the ODEs $\dot{x}_i$, $i = 1, \ldots, 8$, we obtain a mass-action system for the model without intermediates with modified rate constants where $k_i = k_{ib} k_{iy}$ and $k_{ir} = k_{iar} k_{iyr}$. This shows that the rate of formation of phosphoforms in networks with intermediate formation is directly proportional to the strength of the intermediate complex ($\frac{1}{K_m}$) and the rate at which the complex is formed (k_{ia}), depending on the direction of the reaction. This new rate of phosphotransfer (without terms representing variables in the network) indicates that the necessary conditions for sigmoidality and hyperbolic response will remain the same as shown in section (2.2.1.2).

2.3.2 Model with production and degradation

We investigate if the conditions for sigmoidality are altered by the introduction of production and degradation in the model.

Reactions, equations and steady states.
We consider the system with reactions as described in section 2.2.1, together with degradation ($\emptyset$ represents removal of the protein from the network via degradation).

$$\mathrm{HK} \xrightarrow{k_{1o}} \emptyset \qquad \mathrm{HK_p} \xrightarrow{k_{2o}} \emptyset \qquad \mathrm{REC} \xrightarrow{k_{3o}} \emptyset \qquad \mathrm{REC_p} \xrightarrow{k_{4o}} \emptyset$$
$$\mathrm{Hpt} \xrightarrow{k_{5o}} \emptyset \qquad \mathrm{Hpt_p} \xrightarrow{k_{6o}} \emptyset \qquad \mathrm{RR} \xrightarrow{k_{7o}} \emptyset \qquad \mathrm{RR_p} \xrightarrow{k_{8o}} \emptyset$$

and production reactions for the unphosphorylated forms:

$$\emptyset \xrightarrow{k_{1i}} \mathrm{HK} \qquad \emptyset \xrightarrow{k_{3i}} \mathrm{REC} \qquad \emptyset \xrightarrow{k_{5i}} \mathrm{Hpt} \qquad \emptyset \xrightarrow{k_{7i}} \mathrm{RR}$$

We define as usual

$$x_1 = [\mathrm{HK}], \qquad x_2 = [\mathrm{HK_p}], \qquad x_3 = [\mathrm{REC}], \qquad x_4 = [\mathrm{REC_p}],$$
$$x_5 = [\mathrm{Hpt}], \qquad x_6 = [\mathrm{Hpt_p}], \qquad x_7 = [\mathrm{RR}], \qquad x_8 = [\mathrm{RR_p}].$$

The dynamics of the concentrations in time is modelled with a system of ODEs:

$$\dot{x}_1 = -k_s x_1 - k_{2r} x_1 x_4 + k_2 x_2 x_3 - k_{1o} x_1 + k_{1i} \tag{109}$$

$$\dot{x}_2 = k_s x_1 + k_{2r} x_1 x_4 - k_2 x_2 x_3 - k_{2o} x_2 \tag{110}$$

$$\dot{x}_3 = -k_2 x_2 x_3 + k_{h1} x_4 + k_{2r} x_1 x_4 + k_3 x_4 x_5 - k_{3r} x_3 x_6 - k_{3o} x_3 + k_{3i} \tag{111}$$

$$\dot{x}_4 = k_2 x_2 x_3 - k_{h1} x_4 - k_{2r} x_1 x_4 - k_3 x_4 x_5 + k_{3r} x_3 x_6 - k_{4o} x_4 \tag{112}$$

$$\dot{x}_5 = -k_3 x_4 x_5 + k_{3r} x_3 x_6 + k_4 x_6 x_7 - k_{4r} x_5 x_8 - k_{5o} x_5 + k_{5i} \tag{113}$$

$$\dot{x}_6 = k_3 x_4 x_5 - k_{3r} x_3 x_6 - k_4 x_6 x_7 + k_{4r} x_5 x_8 - k_{6o} x_6 \tag{114}$$

$$\dot{x}_7 = -k_4 x_6 x_7 + k_{h2} x_8 + k_{4r} x_5 x_8 - k_{7o} x_7 + k_{7i} \tag{115}$$

$$\dot{x}_8 = k_4 x_6 x_7 - k_{h2} x_8 - k_{4r} x_5 x_8 - k_{8o} x_8. \tag{116}$$

The system does not have any conservation law. Thus, the steady-state equations are given by setting the derivative of the concentration to zero, that is $\dot{x}_i = 0$.

Hyperbolic vs. sigmoidal:

The procedure applied to our initial system in section 2.2.1.1 to obtain the inverse of the signal-response curve, can be applied here to obtain an analytical expression for the inverse of the signal-response curve. The role of the total amounts $\overline{H}, \overline{C}, \overline{T}$ and

$\overline{R}$ is played by the quotients

$$K_1 = \frac{k_{1i}}{k_{1o}}, \qquad K_3 = \frac{k_{3i}}{k_{3o}}, \qquad K_5 = \frac{k_{5i}}{k_{5o}}, \qquad K_7 = \frac{k_{7i}}{k_{7o}}.$$

We do not reproduce the analysis here again. The procedure leads to the derivative of the signal-response curve at zero. Alternatively, we can apply the procedure described in the previous subsection to directly obtain the sign of the second derivative of the signal-response curve at zero, without explicitly computing the signal-response curve. The expression of the second derivative of the signal-response curve at zero is very large, and hence we only provide here the positive monomials with the aim of determining what architectures can exhibit sigmoidality.
We use the definition of K_1, K_3, K_5, K_7 above, together with

$$K_2 = \frac{k_{2o}}{k_{1o}}, \qquad K_4 = \frac{k_{4o}}{k_{3o}}, \qquad K_6 = \frac{k_{6o}}{k_{5o}}, \qquad K_8 = \frac{k_{8o}}{k_{7o}}.$$

The positive terms that can lead to sigmoidality are then

$$S_{o,pos} = k_2K_3K_1(k_{8o} + k_{4r}K_5 + k_{h2})\Big(k_{4r}k_{6o}k_3^2K_5^2K_6(k_2K_3 + k_{2o}) \\ + K_4K_3k_2k_{3r}(k_{h1} + k_{4o})(k_{4r}K_5(k_{3r}K_3 + k_{6o}) + (k_{8o} + k_{h2})(k_4K_7 + k_{3r}K_3 + k_{6o}))\Big).$$

We observe that if $k_{3r} = k_{4r} = 0$ then $S_{o,pos} = 0$ and sigmoidality cannot occur. Contrary to the system without production and degradation, $k_{3r} = 0$ does not guarantee that sigmoidality cannot occur. This is due to the fact that now there is a degradation of Hpt_p, which plays the role of the hydrolysis k_{h1} at REC_p. Therefore, the reverse phosphorelay between layers 3 and 4 can also account for sigmoidality.

In the system with production/degradation reactions, inclusion of intermediates cannot alter steady-state properties such as the existence of sigmoidality. In recent work, we have shown that in reaction schemes that do not give rise to conservation relations, consideration of complex formation does not alter the system properties at steady state [95].

2.3.3 Model with auto-dephosphorylation at HK

We extend the model given in section 2.2.1 to incorporate auto-dephosphorylation of HK. That is, we enrich the system with a reaction

$$\mathrm{HK_p} \xrightarrow{\mathrm{k_5}} \mathrm{HK}.$$

The ODE system modelling the protein concentrations is identical to (26-33) except for the expressions for $\dot{x}_1, \dot{x}_2$ that become

$$\dot{x}_1 = -k_s x_1 - k_{2r} x_1 x_4 + k_2 x_2 x_3 + k_5 x_2$$
$$\dot{x}_2 = k_s x_1 + k_{2r} x_1 x_4 - k_2 x_2 x_3 - k_5 x_2.$$

The steps followed in section 2.2.1 can be applied here as well to obtain an analytical expression for the inverse of the signal-response curve.The sign of the second derivative of the signal-response curve at zero agrees with the sign of

$$S_d := (k_5 + k_2\overline{C})S - k_2 k_5 \overline{H}\big((k_3\overline{T} + k_{2r}\overline{H})k_4 k_{h2} z_3 \overline{R} + z_1 k_3 k_{2r}\overline{CH}(k_{3r} z_2 \overline{C} + k_4 k_{h2}\overline{R}) + k_{h1} z_3^2\big)$$

where

$$z_1 = k_{4r}\overline{T} + k_{h2}, \qquad z_2 = k_4\overline{T} + k_{h2}, \qquad z_3 = k_4 k_{h2}\overline{R} + k_{3r} z_1 \overline{C}$$

and S is given in Eq. (60). We easily see that S_d can only be positive if S is positive. Therefore, the necessary conditions for sigmoidality for the simple model are not altered by explicitly modelling auto-dephosphorylation of HK.

2.3.4 Model with auto-dephosphorylation at Hpt

We extend the model given in section 2.2.1 to incorporate auto-dephosphorylation of Hpt. That is, we enrich the system with a reaction

$$\mathrm{Hpt_p} \xrightarrow{\mathrm{k_5}} \mathrm{Hpt}$$

We applied the steps described in section 2.3.1 to obtain an expression for the sign of the second derivative of the signal-response curve at zero in terms of the rate constants and total amounts. The sign is given by

$$\begin{aligned} S_3 =& S - k_5 z_1 \Big(z_1 z_2 (k_{h1} k_{3r}\overline{C} + k_3 k_5 \overline{T} + (k_{2r}\overline{H} + k_{h1})(k_{3r}\overline{C} + k_5)) + 2\, k_4 k_{h2} z_2 (\overline{H} k_{2r} + z_2)\overline{R} \\ &+ \overline{CH}(k_2 k_3^2 (k_4 - k_{4r})\overline{T}^2 + (k_2 k_3 k_4 - k_2 k_{3r} k_{4r} + k_{2r} k_{3r} k_{4r}) k_{h1}\overline{T} \\ &+ k_{h1} k_{h2}(k_2 k_3 - k_2 k_{3r} + k_{2r} k_{3r})) \Big) \end{aligned} \tag{117}$$

where

$$z_1 = k_{4r}\overline{T} + k_{h2}, \qquad z_2 = k_3\overline{T} + k_{h1}$$

and S is given in eq. (60). The terms highlighted in blue are the terms that can possibly contribute to S_3 being positive. The last two highlighted terms are also highlighted in S in eq. (60) and further, they are multiplied by k_{h1}. When k_5 is set to zero, the sign of the second derivative of the signal-response curve at zero agrees with the corresponding sign for the model without auto-dephosphorylation at Hpt. However, when $k_5 \neq 0$ then sigmoidality can arise even if $k_{h1} = k_{3r} = 0$ but $k_{4r} > k_4$. For this model, necessary conditions for sigmoidality of the signal-response curve are either that

$$k_{h1} > 0 \quad \text{and} \quad k_2 > k_{2r} \quad \text{and} \quad (k_{3r} > k_3 \quad \text{or} \quad k_{3r}k_{4r} > k_3k_4).$$

or that

$$k_5 > 0 \quad \text{and} \quad k_{4r} > k_4.$$

In other words, necessary conditions for sigmoidality of the signal-response curve are

$$k_{h1}(k_2-k_{2r})(k_{3r}-k_3) \neq 0 \quad \text{or} \quad k_{h1}(k_2-k_{2r})(k_{3r}k_{4r}-k_3k_4) \neq 0 \quad \text{or} \quad k_5(k_{4r}-k_4) \neq 0.$$

These equations show the necessary conditions for sigmoidality, which are identical to findings from models without dephosphorylation at Hpt (section 2.2.1). Biologically, it shows that a stronger reverse phosphotransfer along with either a bifunctional phosphatase activity ($k_5 > 0$) or auto-dephosphorylation ($k_{h1} \neq 0$) is essential for sigmoidality.

2.4 Summary of Findings

In this chapter, we built models for a generic four protein phosphorelay and examined the role of different biochemical reactions like reverse phosphorylation, hydrolysis of phosphorylated aspartate residues, total protein concentration etc. on the shape of the overall signal-response relationship observed in the phosphorelays. We did this by generating different variants of the four layered phosphorelay architecture where the position of reverse phosphotransfer and hydrolysis reactions were altered to study all theoretically possible relay structures of length four.

Using a recently developed recursive technique [96–98], we were able to find an analytical description of the steady states of the ordinary differential equations (ODEs) representing different phosphorelay architectures (see section 2.2). For a phosphorelay containing monofunctional histidine kinase, we managed to derive an explicit analytical description of the signal-response curve. Analysis of the

relations among the concentrations at steady state revealed that 14 of the 32 possible topologies result in non-responsive systems, where the level of phosphorylated RR reaches its maximum for any non-zero signal (see Appendix B.1.1). Common to all the 14 non-responsive topologies is the absence of a hydrolysis reaction on RR ($k_{h2} = 0$) and, additionally, either there is no hydrolysis reaction on REC ($k_{h1} = 0$), or there are no reverse phosphorylation reactions between REC-Hpt ($k_{3r} = 0$) or between Hpt-RR ($k_{4r} = 0$). In other words, the ability of a four-layered phosphorelay to respond to a range of signals necessitates the presence of either hydrolysis from RR, or both hydrolysis from REC and reverse phosphorylation at one of the final two layers.

For the remaining 18 responsive topologies we sampled 1000 parameter sets (rate constants and total protein concentrations) from a biologically permissible range, derived the signal-response curve for each parameter set and classified this curve as hyperbolic or sigmoidal (see sections 2.2.1 and 2.2.1.3). The hyperbolic case contains linear signal-response relationships with saturation. The sigmoidal case indicates that the signal-response relationship includes an inflection point [40], and could endow the cell with switch-like responses and decision-making [99] via the phosphorelay. A sigmoidal signal-response relationship can also embed ultrasensitivity [40, 100]. The classification of the signal-response curves resulting from parameter sampling revealed that out of the 18 topologies, only 4 allowed sigmoidality in any significant part (more than 2%) of the sampled parameter space and when considering both equal and different total protein concentrations at different layers (see table 2.4, figure 2.4 & figure 2.5).

Interestingly, common to all these topologies is a presence of reverse phosphorylation between REC-Hpt and between Hpt-RR, as observed in nature [3]. Of the 4 topologies, only two (topologies labelled 14 and 30) resulted in an equal distribution of sigmoidal and hyperbolic responses among the sampled parameter sets suggesting that their signal-response relationship can easily be tuned (table 2.4). One of them (topology 30) represents the configuration that is observed in a large number of four protein phosphorelays studied to date with reverse phosphorylation at layers three and four, combined with hydrolysis from layers two and four [29, 30, 101–103].

To further understand the effect of reverse phosphorylation in generating sigmoidality, we compared the sampled parameter sets resulting in hyperbolic vs. sigmoidal signal-response relationships. We found that a key difference between the two parameter sets is the ratio between the forward and reverse phosphorylation rate constants, where a mean ratio below one is observed in the case of sigmoidal

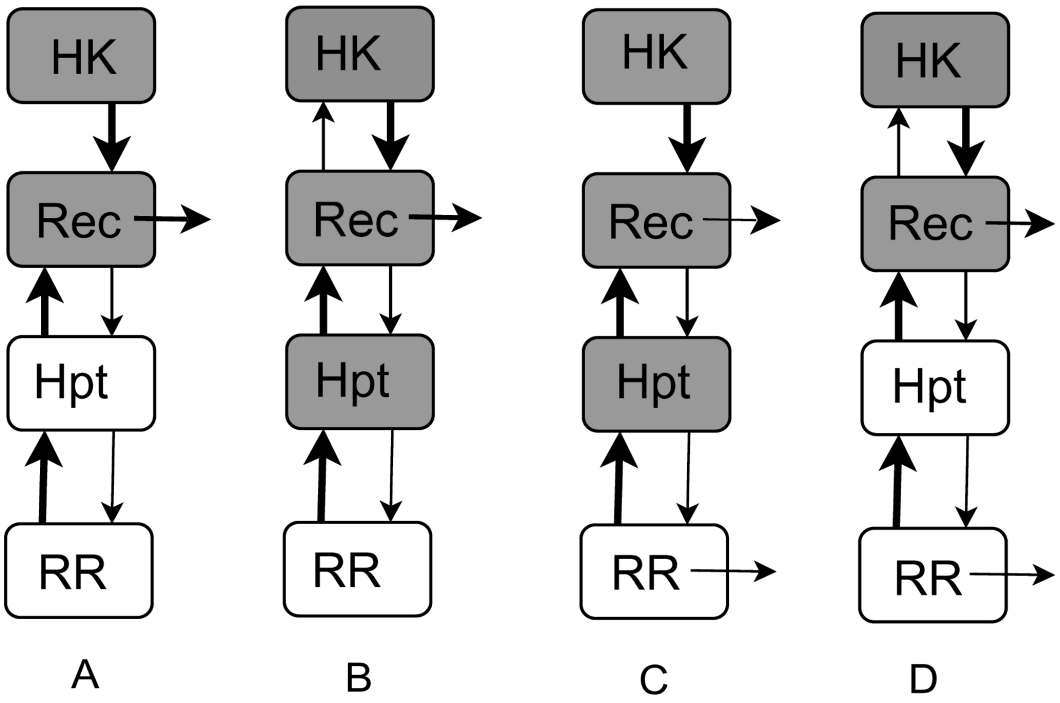


Figure 2.4: Schematic representations of the four topologies that allowed sigmoidal signal-response relationships in a significant part (more than 2%) of the sampled parameter space when considering monofunctional HK and different total protein concentrations at different layers (see table 2.4). Panels A, B, C and D show topologies 14, 16, 30 and 32 respectively, each corresponding to a specific set of reverse phosphotransfer and hydrolysis reactions being present. Reactions are shown as directional arrows, where thickness of the arrow indicates the relative strength of the reaction. In other words, arrows are weighted by the mean reaction rate constant obtained from all sampled parameter sets producing sigmoidality. For each layer and a given topology, a grey (open) backdrop indicates that the mean of total protein concentration at that layer is high (low), based on all sampled parameter sets producing sigmoidal signal-response curves.

signal-response relationships, , while in hyperbolic signal-response relationships, the ratio is greater than one (see table 2.8, figures 2.4 & figure 2.5). To analytically confirm if the reverse phosphorylation rate constant being higher than the forward phosphorylation rate constant is a necessary condition for achieving sigmoidality in phosphorelays, we computed analytically the second derivative of the signal-response relationship (section 2.2.1). Note that a hyperbolic curve has negative second derivative throughout its domain (in our case, for positive signals), while the second derivative of a sigmoidal curve is initially positive and then it changes sign. Thus, the sign of the second derivative of the signal-response curve at zero can be taken as a test for sigmoidality. Using the analytical description of the second derivative of the signal-response curve at zero we found three necessary analytical conditions for achieving a sigmoidal signal-response relationship: (i) $k_{h1} > 0$, (ii) $k_2 > k_{2r}$ and (iii) $k_{3r} > k_3$ or $k_{3r}k_{4r} > k_3k_4$, where k_{h1} is the hydrolysis rate for REC and k_2 (k_{2r}), k_3 (k_{3r}) and k_4 (k_{4r}) are the rate constants of forward (reverse) phosphotransfer

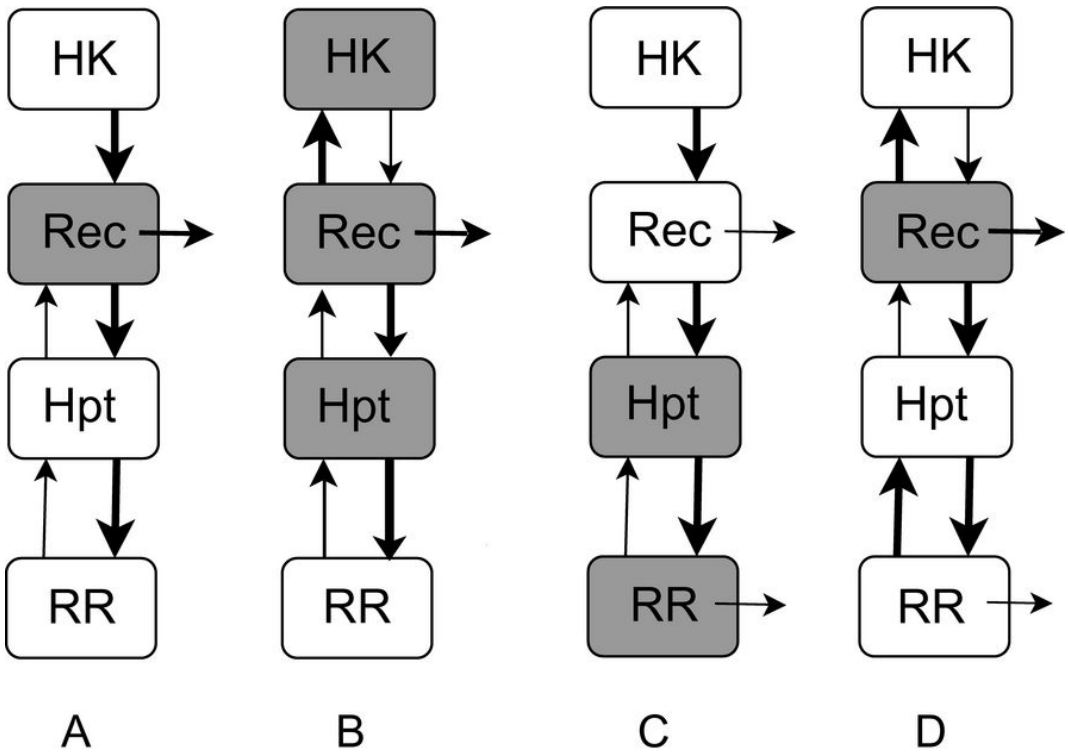


Figure 2.5: Schematic representations of the four topologies 14, 16, 30 & 32 in a hyperbolic signal-response regime. Arrows are weighted by the mean reaction rate constant obtained from all sampled parameter sets producing hyperbolic signal-response curves. For each layer and a given topology, a grey (open) backdrop indicates that the mean of total protein concentration at that layer is high (low), based on all sampled parameter sets producing hyperbolic signal-response curves. Panels A, B, C and D show topologies 14, 16, 30 and 32 respectively, each corresponding to a specific set of reverse phosphotransfer and hydrolysis reactions being present. Reactions are shown as directional arrows, where thickness of the arrow indicates the relative strength of the reaction.

reactions between HK-REC, REC-Hpt and Hpt-RR respectively. If either (i), (ii) or (iii) are not fulfilled, then the signal-response curve is hyperbolic. As a consequence, k_{h1} and k_{3r} are required to be non-zero for sigmoidality to occur. This analytical result is in full agreement with the classification of signal-response curves resulting from parameter sampling. More particularly, the above mathematical conditions explain why only topologies embedding hydrolysis at REC and reverse phosphotransfer between REC-Hpt display sigmoidality, and why only topologies where these reactions are coupled with reverse phosphotransfer between Hpt-RR result in sigmoidality in a larger portion of the parameter space (see table 2.4 and figure 2.4). We further conclude that the rate constant of the reverse phosphotransfer at HK-REC must be small for sigmoidality to arise (see section 2.2.1).

These results can be understood intuitively if we consider the phosphorelay as a set of connected stations, through which phosphoryl groups flow at a rate dictated by

	Topology 14		Topology 30	
Parameter Ratio	Hyperbolic Regime	Sigmoidal Regime	Hyperbolic Regime	Sigmoidal Regime
$\frac{k_3}{k_{3r}}$	3.731	0.8	5.552	0.831
$\frac{k_4}{k_{4r}}$	3.329	0.886	8.66	0.842

Table 2.8: The mean of the ratio of forward to reverse rate constants based on samples resulting in hyperbolic and sigmoidal signal-response relationships. The results shown are for topologies 14 and 30, assuming monofunctional HK, and using classifications based on the second derivative of the signal-response curve at zero and from sampling all parameters (with equal total protein concentrations at different layers). For additional results using alternative classification and sampling schemes (different total protein concentrations at different layers), assuming bifunctional HK, as well as for results from topologies 16 and 32, see Appendix A.

the signal strength. Without the presence of reverse phosphorylation and hydrolysis reactions in intermediate layers, phosphoryl groups accumulate at the bottom of the relay at a constant rate, while intermediate layers can remain unphosphorylated until the layers below them are saturated [33, 96]. When hydrolysis from the bottom layer is absent, this saturation effect becomes immediate for the last layer, creating a non-responsive system (as discussed above). Reverse phosphotransfer reactions at layers 3 and 4 generate a back-flow from the layer that they are embedded at, thereby increasing the signal level required for the saturation of the bottom layer with phosphoryl groups. This buffering effect presents itself in the signal-response curve of a given layer as sigmoidality, where the phosphorylated form in that layer can remain at low levels despite high signal flow from the top of the relay. It can be expected that implementation of subsequent reverse phosphotransfer and hydrolysis reactions from a given layer would increase the buffering effect and result in higher levels of sigmoidality in the signal-response relationship. This intuitive picture is in line with the analytical results described above, which reveal k_{h1}, k_2, $k_{3\mathrm{r}}$ and $k_{4\mathrm{r}}$ as the key parameters that control the shape of the signal-response relationship (i.e. its sigmoidality) for the last layer (section 2.2.1). It should also be noted that sigmoidality at the last layer of a phosphorelay could still be achieved in relays of shorter or longer length, provided that the general principles outlined above are met through the use of reverse (or cross) phosphotransfer and hydrolysis reactions.

The analyses described so far have several assumptions with regards to modeling phosphorelay dynamics. Firstly, we have assumed bimolecular phosphotransfer reactions without complex formation. This assumption would be satisfied if phosphotransfer reactions, which are distinct from enzyme-driven reactions, happen fast and any complexes formed are short-lived. While there is some indication from *in vitro* phosphotransfer reactions that this might be the case (e.g. [104]), we

have relaxed the assumption of no complex formation and developed a model that includes complex formation at each layer of the relay. By suitable identification of the rate constants of the system without intermediates to the rate constants of the system with intermediates (see section 2.3.1), we show that the originally identified condition - that reverse phosphorylation between layers 2-3, and hydrolysis are necessary for sigmoidality - also holds when considering complex formation.

Secondly, we have assumed constant total protein concentrations in each layer, ignoring the effects of any processes such as expression, degradation and dilution. This assumption would be valid if such processes happen at much slower time scales compared to the signalling dynamics of the relay. Relaxing this assumption and considering production/degradation processes as simple in and out fluxes for un-phosphorylated and phosphorylated proteins respectively, we derived analytically, an expression for the second derivative of the signal-response curve at zero (see section 2.3.2). Necessary conditions for sigmoidality in this system are either that k_{3r} (reverse phosphotransfer from Hpt to REC) or that k_{4r} (reverse phosphotransfer from $\mathrm{RR_p}$ to Hpt) is non-zero. These conditions differ from the necessary conditions for sigmoidality in the simple model. The main differences are that hydrolysis at the second layer (i.e. $k_{h1} > 0$) is no longer a required condition, and that k_{3r} can be zero as long as k_{4r} is not. The first difference arises because the degradation reaction of phosphorylated REC mimics the role of k_{h1}. Similarly, the second difference is due to the degradation reaction of phosphorylated Hpt and RR in the third and fourth layers. These equalise the roles of the reverse phosphotransfer reactions at the third and fourth layers in controlling the shape of the signal-response curve.

Thirdly, we have assumed the absence of auto-dephosphorylation of HK and Hpt, as these proteins get phosphorylated on a histidine residue which is indicated to be stable against auto-dephosphorylation. We find that explicitly modelling the auto-dephosphorylation of HK does not alter the conclusions with regards to the necessary conditions for sigmoidality (section 2.3.3). When assuming auto-dephosphorylation of Hpt, we find that the necessary conditions required for sigmoidality are either that the necessary conditions for the simple model hold, or that $k_{4r} > k_4$ (see section 2.3.4). As a consequence, sigmoidality can arise even if k_{h1} and k_{3r} are zero, that is, in the absence of hydrolysis at REC and reverse phosphotransfer between Hpt and REC. This can be explained similarly to why degradation reactions alter the necessary conditions for sigmoidality in the simple model.

In the above treatment, we have also assumed a monofunctional HK, while it is

known that several HKs can show both phosphorylation and dephosphorylation activity towards their substrate (in this case REC). We find that considering such a bifunctional HK does not alter the overall analytical conclusions regarding the necessity of fast reverse phosphotransfer and presence of hydrolysis reactions for enabling sigmoidality in the system (see Appendix C). We find that the addition of a bifunctional HK can have significant effects on the distribution of the signal-response relationship classification in specific topologies (see tables 2.6 & 2.7). In particular, topologies with reverse phosphotransfer reactions in all layers exhibit a drop in the number of parameter sets showing sigmoidality when HK is bifunctional, while topologies lacking hydrolysis at REC can exhibit sigmoidality where they could not under the simple model.

The latter finding is understandable as the bifunctional HK-mediated dephosphorylation can mimic the effects of hydrolysis at REC (k_{h1}). We found that varying the rate constant of the HK-mediated dephosphorylation at REC (k_5) in comparison to changing k_{h1} has similar but stronger effects on the shape of the signal-response relationship (see figure 2.6). The additional dephosphorylation reaction mediated by the bifunctional HK alters the analytical description of the signal-response relationship in such a way that several previously non-responsive topologies become responsive, while the maximal level of phosphorylated RR at steady state remains unaltered (section 2.2.3). However, due to sequestration of phosphorylated REC by the bifunctional HK and subsequent dephosphorylation, a higher signal level is required when HK is bifunctional to achieve the same level of response as in the monofunctional case (for a given set of parameters).

To understand the consequences of hyperbolic vs. sigmoidal signal-response relationships in a phosphorelay, we focused on the two topologies that displayed high levels of tunability between these two response types (topologies 14 and 30) and further analyzed the signal-response relationship. As explained above, both of these topologies embed reverse phosphotransfer reactions between REC-Hpt and between Hpt-RR. They differ, however, in the implementation of hydrolysis reactions; topology 30 embeds hydrolysis at the level of both REC and RR, while topology 14 embeds hydrolysis only at the level of REC. For each topology we picked 100 random parameter sets from both hyperbolic and sigmoidal regimes (i.e. parameters resulting in hyperbolic and sigmoidal signal-response relationships), and analyzed the noise properties and response time of the resulting systems (section 2.2.2.2). To understand the reliability of the response, we examine the network's noise in output i.e. we are measuring the standard deviation in the response (RR_p) levels with respect to different input (k_s) levels. In this study, we did not

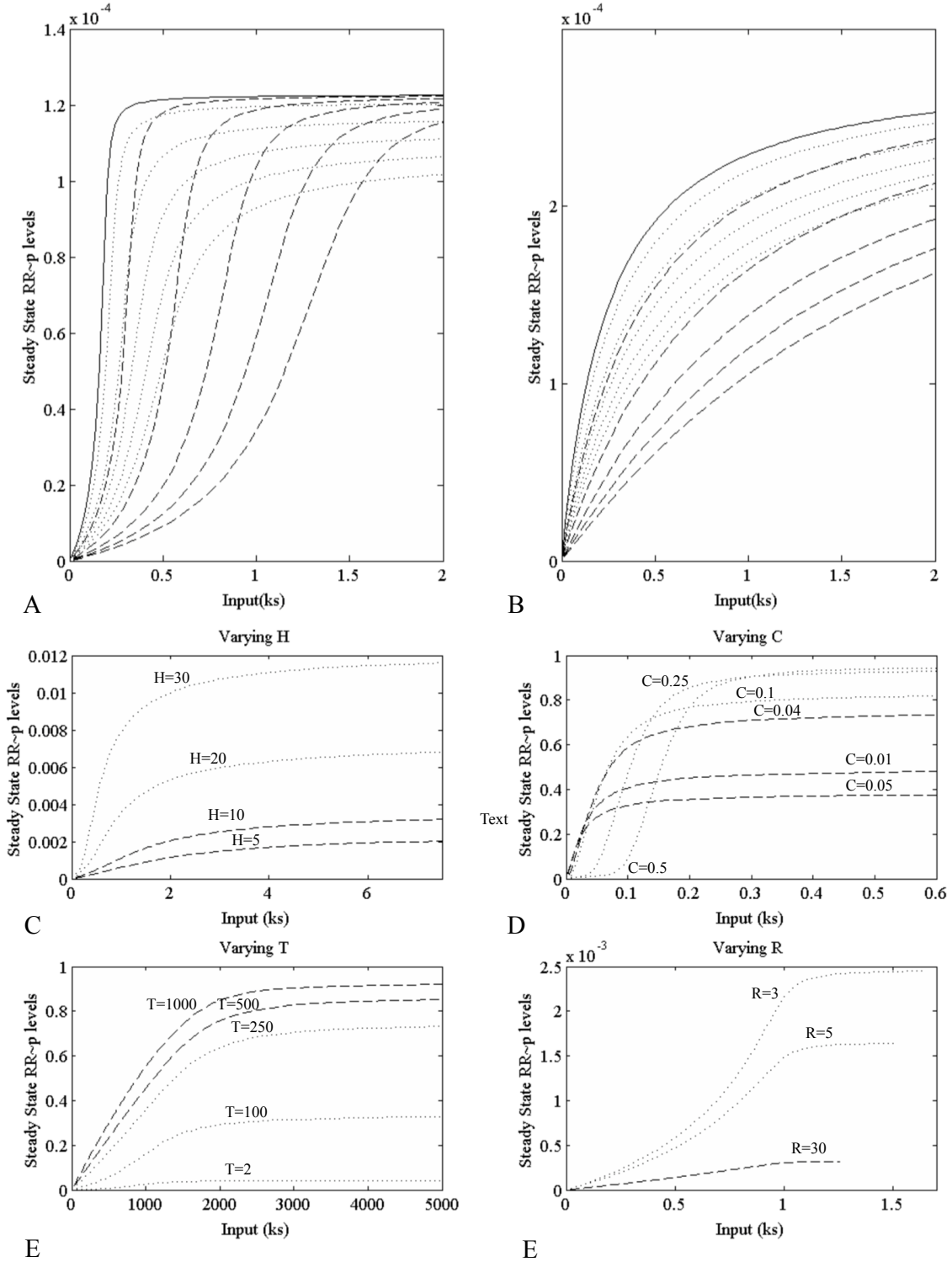


Figure 2.6: Panels (A & B) show the effects of varying k_5 (dashed line) and k_{h1} (dotted line) on the signal-response curves in both the sigmoidal and hyperbolic regimes. Panels (C, D, E and F) show that the shape of the signal-response curve can be tuned from one regime to another by varying total protein levels at different layers of the relay. The x-axis is the signal to the system (the HK auto-phosphorylation rate constant, k_s), while the y-axis corresponds to the concentration of phosphorylated RR. Each line represents a system with varying total protein levels. Parameters used to generate the curves in this figure are described in appendix A.2.

introduce noise from outside the network but instead used PRISM to measure the reliability of the network's response when reactions in the network are modelled as probabilistic processes i.e. we examined the intrinsic noise or variations in the network. These variations are due to the probabilistic nature of chemical reactions where collisions between reactants results in product formation, and therefore is dependent on the probability of reactant molecules colliding with each other. We find that for both topologies, the phosphorylated RR levels across the full signal range displayed higher levels of noise in the sigmoidal regime compared to the hyperbolic regime (figure 2.2). This finding is also in line with previous theoretical findings, which showed that the level of noise in a dynamical system is proportional to the level of signal amplification it implements [104]. The finding also fits the experimental results of the *B. subtilis* sporulation phosphorelay, which embeds topology 30 discussed here, and displays high levels of heterogeneity in its output [36].

In contrast to the results from the noise analysis, the results of the response time analysis differed for the two topologies. Response time refers to the time required for the phosphorylated RR levels to reach steady state following a step increase (or drop) in signal levels (section 2.2.2.1). With the parameters set to result in sigmoidality, topology 30 displayed a smaller response time compared to the case when parameters were set to result in a hyperbolic signal-response relationship (see figures 2.2 & 2.3). For topology 14, however, we did not observe any difference in response times under hyperbolic and sigmoidal regimes. These results can be understood in light of the different hydrolysis reactions present in these two topologies. Having both hydrolysis reactions present, topology 30 displays a lower phosphorylated RR level at saturation and requires a higher signal level to reach the same level compared to topology 14 (compare figure 2.2A & figure 2.2B). At the same time, however, the presence of both hydrolysis reactions in topology 30 could allow more parameters for tuning the level of sigmoidality and also enhance the effect of the difference between reverse and forward phosphorylation rates from the third layer (see eq. (60)). This could provide the basis for the observed improvement in response time in topology 30, i.e. the ability to achieve sigmoidality via less pronounced reverse phosphorylation rates between Hpt-RR, such that the response at the level of RR is not slowed down.

2.5 Discussion & Conclusions

In this chapter, we undertook an analytical and simulation based study to decipher the functional role of the conserved features of these phosphorelays. The findings demonstrated that either the presence of hydrolysis from RR, hydrolysis from REC or reverse phosphorylation between REC-Hpt or Hpt-RR are necessary to achieve

a responsive four-layered phosphorelay. These structural constraints to achieve a functional signalling system are further refined if the signal-response relationship of such a relay is to be sigmoidal. In particular, we prove that necessary conditions for sigmoidality are the presence of hydrolysis at the second layer, high forward flow of phosphoryl groups at the second layer, and high reverse flow of phosphoryl groups at the third and final layers. The last condition directly controls the shape of the signal-response relationship, with a ratio of forward to reverse phosphorylation rate constants above (below) one strongly favouring a sigmoidal (hyperbolic) relationship. The noise characteristics and response times are different in the two regimes resulting in noisier and faster signalling from the phosphorelay when it operates in the sigmoidal regime. We find that bifunctionality of a HK does not alter substantially these conclusions.

These results provide mathematical proof that the way in which reverse phosphorylation in phosphorelays is implemented in natural systems endows functionality and allows tuning of signal-response relationships between a hyperbolic and sigmoidal regime. Together with previous mathematical analyses of phosphorelays, which showed that the maximal level of phosphorylated RR and the signal-to-noise ratio of the response saturate at a relay length of four [33, 96], these findings provide a possible explanation for observed phosphorelay structures. It is plausible that relay length and specific location of reverse phosphorylation and hydrolysis reactions have evolved towards achieving signal processing capability. Evolution could have then exploited specific regimes of rate constants to achieve higher plasticity in the signal-response relationship phosphorelays could embed. In particular, the naturally observed structure makes it possible to tune the signal-response relationship of a phosphorelay both through genetic mutations affecting reaction rate constants and through regulatory interactions. The latter could include regulating the total protein concentrations at the different layers of the relay (e.g. via transcriptional regulation or feedback), altering reaction rate constants through binding of auxiliary proteins on relay components, and regulating the bi-functional activity of a HK. Indeed, we find that all these parameters have a significant effect on the shape of the signal-response relationship (figures 2.2 and 2.6). There is some empirical evidence that cells might be exploiting such alterations as regulatory points. For example, the kinase and phosphatase activities of certain bifunctional HKs may be regulated through binding of auxiliary ligands [17, 105] or two-component proteins [106]. Experimental studies towards exploring the presence and extent of the other possibilities should be facilitated by the presented results.

All these finding are in line with the observations from the *B. subtilis* sporulation

phosphorelay, which is the phosphorelay with the most extensive characterization of response dynamics. This relay displays topology 30 presented here, and has been recently shown to display ultrasensitivity [37] and high levels of noise [36]. It has been so far not clear where the source of such ultrasensitivity lies [37]. The presented results suggest that the relay dynamics itself could be the primary source of these observed dynamics and noise. Indeed, we find that the *in vitro* measured kinetic rates from this relay ([55]) fit the necessary conditions we derived here for sigmoidality. More intriguingly, the *B. subtilis* relay features transcriptional feedbacks to the second and fourth layers. As discussed above, such regulation on the concentrations of relay components could allow tuning of the signal-response relationship between hyperbolic and sigmoidal (see figure 2.6). This could contribute to the observed complexity in the dynamics of this system [3, 36, 37]. Thus, considering relay dynamics in light of the findings presented here could help design future experiments to better understand the control of the sporulation decision in *B. subtilis*. More generally, the presented findings allow for classification of the signal-response relationship of a phosphorelay from *in vitro* constitution of its parts and measurement of the specific phosphotransfer reaction rates using radiolabelled phosphate groups. Such *in vitro* measurements are commonly employed in the study of bacterial two-component systems [55, 107, 108], and while they cannot be entirely conclusive about the *in vivo* rates, provide an insight about the parameter regime in which the kinetics of a phosphotransfer reaction resides. Combined with the findings presented here, such measurements would be informative for further experimental designs (e.g. analysis of population level heterogeneity would be interesting to pursue if signalling network dynamics indicates sigmoidal signal-response relationships).

In summary, the main conclusions of this part of the study are that phosphorelays can embed hyperbolic or sigmoidal signal-response relationships, and that the latter type is not possible without reverse phosphorylation and a hydrolysis reaction at the second layer. Achieved either via dynamical tuning or through evolution of kinetic rates, the hyperbolic and sigmoidal regimes should allow appropriate physiological responses as needed by the cell. We would expect that sigmoidal dynamics would be favoured when binary decision making is required. In contrast, hyperbolic or linear signal-response relationships would be required to produce responses that track the incoming signals. Classifying a given phosphorelay's behaviour into these regimes would be highly valuable, but is currently hampered as measuring the response of a phosphorelay at different signal levels and/or different component concentrations is highly difficult. Further, the signals feeding into phosphorelays are often unknown or not feasible for experimental manipulation. The results presented here offer an alternative, in which the shape of the signal-response relationship of the relay can

be predicted from the measurement of forward and reverse phosphorylation rates. These measurements are possible in most cases through *in vitro* phosphotransfer experiments, as discussed above for the *B.subtilis* system, and hence can provide a direct prediction of the *in vivo* signalling dynamics that can be further tested.

Mutations and gene duplications provide the mechanisms by which the structure and dynamics of cellular interaction networks can be changed in evolution. Mathematical and computational approaches such as the ones presented here allow mapping the signal-response relationship of the possible systems that can be generated in this way. This understanding is essential to grasp why evolution might have resulted in the observed features of biological systems and how we might further modulate them. Thus, our findings on phosphorelays should facilitate both understanding the physiology mediated by these systems in a wide range of organisms and (re)engineering these through synthetic biology.

3

Unlimited multistability in microbial signalling networks

Chapter Overview

Building on the detailed examination of different biochemical reactions occurring in different phosphorelay architectures, we explore the dynamics of networks containing multidomain sensor histidine kinase proteins and compare it with networks which have sensor kinases where there is only one phosphotransfer domain. This is important as unlimited multistability is known to arise from multi-site phosphorylation seen in the signalling networks of eukaryotic cells, but a similar universal mechanism has not been identified in microbial signalling networks. This chapter explores the effects of multidomain proteins in prokaryotic signalling networks and identifies how interconnected feedback loops amongst different phosphoforms of these multidomain sensor histidine kinases tied to a downstream substrate give rise to multistable signal response dynamics.

After identifying the core sub network in the overall signalling architecture that implements multistability, we examine how the dynamics of sharing proteins

and combining different inputs with a common response depends on the number of shared proteins and its position in a phosphorelay. We find that there is unbounded multistability with a direct co-relation between the number of steady states and n sites of phosphorylation. By testing simple architectures with two different sensor histidine kinases sharing common phosphotransfer proteins acting on a common response regulator, we examine how combining multiple inputs via shared components has a potential for implementing Boolean logic in some network architectures. By sampling within biologically acceptable parameter regimes, we identify parameters where these architectures with shared components implement AND, OR and ADDER functions in cells. Using two experimentally studied examples of two-component systems implementing hybrid HKs, we show that bistability and implementation of logic functions are possible under biologically feasible reaction rates. Microbial genomes contain significant numbers of hybrid and unorthodox HKs, with some genomes having a larger fraction of these proteins compared to regular HKs. We show that microbial cells are thus theoretically unbounded in mapping distinct environmental signals onto distinct physiological states and performing complex computations on them.

Finally, we will discuss how findings from some of these simple theoretical architectures are applicable to larger and more complex naturally occurring systems in a cell and how these findings facilitate the understanding of natural two-component systems and allow their engineering through synthetic biology.

3.1 Introduction

The ability to map environmental signals onto distinct internal physiological states or programmes is critical for single-celled microbes. This ability requires mapping different environmental signals, or combinations thereof, onto specific physiological responses in a reliable fashion. A crucial systems dynamics feature underpinning such an ability is multistability. Prokaryotic signalling systems contain two-component systems comprising histidine kinase (HK) receptors and response regulator proteins engaging in phosphotransfer reactions. Understanding the basis of this ability from the viewpoint of systems dynamics, as well as biochemical implementations, is thus crucial for the understanding of cellular behaviour in systems biology and its re-engineering in synthetic biology.

From a systems dynamics perspective, multistable cellular systems such as signalling networks can display abrupt transitions among different steady states when changes

in specific system parameters cross threshold points [39]. Furthermore, the threshold dynamics under multistability can allow cells to generate binary responses to environmental signals, thereby providing the potential for implementing Boolean logic [109]. This threshold dynamics is the hallmark of multistability and is observed in several cellular responses including the all-or-none type responses seen in eukaryotic cell fate determination [99] and cell cycle regulation [110] and is indicated to underpin cellular differentiation [111].

From a mechanistic viewpoint, a key question is how multistability can be implemented through biochemical reactions. Answering this question could allow us to link observed biochemical features of natural systems to higher level response dynamics and exploit certain biochemistries to engineer cell behaviour. There has already been significant progress in both directions, with transcriptional feedback [111, 112] and multi-site phosphorylation [113, 114] identified as key biochemical mechanisms for implementing multistability. These mechanisms are found commonly in nature and have already been exploited in synthetic biology to engineer bistable gene expression and ultrasensitive signal processing [111, 112, 115–117]. In particular, multi-site phosphorylation is proposed as a very general mechanism to generate unbounded multistability [118, 119]. It has been mathematically proven that a protein with n phosphorylation sites catalysed by enzymes in a distributive, sequential manner can give rise to at least $n+1$ steady states [118, 119]. Subsequent theoretical studies show that a sharing of enzymes (i.e. kinases and phosphatases) among different phosphorylation steps and the linking of these steps are crucial prerequisites for multistability in a multi-site phosphorylation system [97, 120].

Interestingly, multi-site, enzyme-mediated phosphorylation as seen in eukaryotic systems is mostly lacking in microbes. Instead, microbes rely on the so-called two-component systems for their environmental sensing and inter-cellular signalling [11]. Biochemically, two-component signalling is very distinct from enzyme-mediated phosphorylation dominating eukaryotic signalling and relies on phosphotransfer reactions between histidine and aspartate residues on histidine kinases (HKs) and response regulator (RR) proteins [11]. Since this biochemistry precludes the enzyme-mediated mechanisms of multistability generation [119] and phosphotransfer occurs via collisions between unphosphorylated proteins and phosphate carrying biomolecules in the absence of an enzyme, it raises the question of whether microbes use a different mechanism for generating multistability or lack this feature altogether. Although specific biochemical arrangements in some two-component systems are shown to enable bistability [35, 121, 122] and several microbial phenotypes are indicated to exhibit bistability [123, 124], a general mathematical framework for

assessing the capacity of system dynamics in two-component signalling has been lacking. Here, we develop such a framework and particularly consider the system dynamics arising from multidomain sensor HKs in prokaryotic signalling.

We find that the presence of these multidomain proteins can allow the system to display bistability, where systems with regular (also called classic or regular) HKs cannot. We show that bistability arises from, and necessitates, the reactions among the different phosphorylation states of the multidomain HK and a downstream protein. Extending from this result, we provide a mathematical proof to show that n multidomain HKs sharing the same downstream component can result in a multistable system with $3n$ steady states. We find that this system dynamics property is easily used to implement Boolean logic using multidomain HKs sensing different signals. Finally, we find that two experimentally studied systems, found in yeast osmoregulation and *Vibrio harveyi* quorum sensing, employ hybrid HKs and display a capacity to implement logic functions and bistability with hysteresis as expected by the presented theoretical framework. We begin with an introduction to different types of multidomain proteins found in prokaryotic signalling networks.

3.1.1 Multi Domain Histidine Kinase and their phosphoforms

Two-component signalling networks and phosphorelays containing different two component proteins [11] are present across all studied microbial genomes to date, with some environmental bacteria showing more than 60 distinct two-component systems [81, 125]. In some phosphorelays the four stages of phosphotransfer can be encoded on separate proteins as seen for example in the phosphorelay regulating *Bacillus subtilis* sporulation decision [3], or fewer proteins with multidomain sensor kinase proteins where REC and Hpt domains can be embedded on a single protein known as a hybrid HK or form an unorthodox HK (embedding both REC and Hpt domains) [30] (figure 1.2). All three types of HKs, regular, hybrid and unorthodox, are found to coexist in many microbial genomes, as well as in plants [28, 30, 126].

With multidomain HKs having more than one phosphotransfer domain, they embed multiple phosphotransfer reactions on the same protein, where each domain can be either phosphorylated (P) or unphosphorylated (O). Unlike regular HKs which have only two states: phosphorylated or unphosphorylated, depending on the position of phosphate groups, hybrid and unorthodox HKs have $2^2 = 4$ and $2^3 = 8$ phosphoforms (see figure 1.2) respectively.

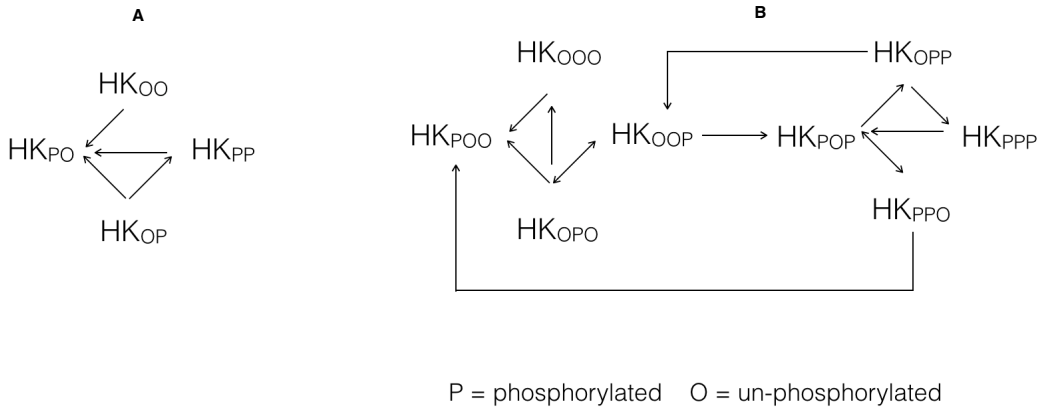


Figure 3.1: Representative figure showing all possible transitions amongst different phosphoforms of HK. A hybrid HK has two phosphotransfer domains i.e. $2^2 = 4$ phosphoforms (HK_{OO}, HK_{PO}, HK_{OP} and HK_{PP}) (A) and an unorthodox HK with 3 phosphotransfer domains has $2^3 = 8$ phosphoforms (HK_{OOO}, HK_{POO}, HK_{OPO}, HK_{OOP}, HK_{POP}, HK_{OPP}, HK_{PPP} and HK_{PPO}) (B). Arrows joining two different HK forms represent a biochemical transition reaction that occurs between the two connected HK forms, with the arrow head representing the direction of phosphotransfer.

In chapter 2, we carried out a thorough exploration of different four protein phosphorelay architectures, and examine the role that different biochemical reactions in the network play in defining specific response dynamics to a phosphorelay. In this chapter, we will focus on phosphorelays with different multidomain sensor proteins and test if this leads to a reaction structure with a capacity for multistability in the network.

3.2 Full, core, and shared components in multidomain HK mediated phosphorelays

3.2.1 Full Hybrid HK mediated Phosphorelay

Phosphotransfer reactions in a hybrid HK mediated phosphorelay are similar to the reactions seen in section 2.2.1, excluding those reactions which involve REC. As the REC domain in a HK mediated phosphorelay is found on the same protein as the histidine kinase, phosphate groups arriving at the HK travel via an intramolecular phosphotransfer from HK to REC before reaching Hpt and RR. This results in the occurrence of HK in different phosphoforms (see figure 3.1). The resulting full set of phosphotransfer reactions are shown in reactions (118)-(124).

$$HK_{OO} \xrightarrow{k_1} HK_{PO} \xrightarrow{k_2} HK_{OP} \xrightarrow{k_3} HK_{PP} \tag{118}$$

$$\mathrm{HK_{OP}} + \mathrm{Hpt} \underset{k_7}{\overset{k_4}{\rightleftharpoons}} \mathrm{HK_{OO}} + \mathrm{Hpt_p} \tag{119}$$

$$\mathrm{HK_{PP}} + \mathrm{Hpt} \underset{k_8}{\overset{k_5}{\rightleftharpoons}} \mathrm{HK_{PO}} + \mathrm{Hpt_p} \tag{120}$$

$$\mathrm{HK_{OP}} \overset{k_9}{\rightarrow} \mathrm{HK_{OO}} \tag{121}$$

$$\mathrm{HK_{PP}} \overset{k_{10}}{\rightarrow} \mathrm{HK_{PO}} \tag{122}$$

$$\mathrm{Hpt_p} + \mathrm{RR} \underset{k_{12}}{\overset{k_{11}}{\rightleftharpoons}} \mathrm{Hpt} + \mathrm{RR_p} \tag{123}$$

$$\mathrm{RR_p} \overset{k_{13}}{\rightarrow} \mathrm{RR} \tag{124}$$

These reactions are initially examined using the CRNT Toolbox ([127–129]). We find that the underlying chemical reaction network (CRN) has the capacity for bistability (see section 1.2.6). To identify the reactions and proteins responsible for this bistability, we iteratively remove reactions from the larger reaction network described in reactions (118) -(124) until we derive a minimal core network which is bistable .[1] This minimal core network contains a hybrid HK occurring as different phosphoforms ($\mathrm{HK_{OO}}$, $\mathrm{HK_{PO}}$, $\mathrm{HK_{OP}}$ and $\mathrm{HK_{PP}}$) along with Hpt (as Hpt & $\mathrm{Hpt_p}$). The reactions in this core include forward phosphotransfer (reactions involving the rates $k_1 \ldots k_5$ in (118)-(124)) and hydrolysis reactions at $\mathrm{Hpt_p}$ (125).

$$\mathrm{Hpt_p} \overset{k_6}{\rightarrow} \mathrm{Hpt} \tag{125}$$

In the original scheme represented by reactions (118)-(124), there is no hydrolysis reaction at $\mathrm{Hpt_p}$. But in the core hybrid network, we introduce reaction (125) to ensure that phosphate groups are removed from $\mathrm{Hpt_p}$ and the core network generates a non-constant steady state response represented by the concentration of $\mathrm{Hpt_p}$ in the phosphorelay. We find that the core bistable network does not require any reverse phosphotransfer (the reaction represented by k_7 in (119)) and is therefore not one of the necessary mechanisms responsible for introducing bistability. In summary, the minimal core network shows that bistability in hybrid HK mediated phosphorelays is due to transitions amongst different phosphoforms of HK, Hpt and $\mathrm{Hpt_p}$.

To understand this better and identify necessary and sufficient conditions causing the

[1] We define a minimal bi(multi)stable network as one where any further removal of proteins or reactions in the network results in the loss of bi(multi)stability.

observed bistability in the network, we derive an ODE model representing reactions in the bistable core and try to arrive at analytical expressions for variables in the system.

3.2.2 Modelling the core bistable network in a hybrid HK mediated phosphorelay

Reactions occurring in a bistable minimal core network are:

$$\mathrm{HK}_{00} \xrightarrow{k_1} \mathrm{HK}_{\mathrm{P}0} \xrightarrow{k_2} \mathrm{HK}_{0\mathrm{P}} \xrightarrow{k_3} \mathrm{HK}_{\mathrm{PP}} \qquad \mathrm{HK}_{0\mathrm{P}} + \mathrm{Hpt} \xrightarrow{k_4} \mathrm{HK}_{00} + \mathrm{Hpt}_{\mathrm{P}}$$

$$\mathrm{Hpt}_{\mathrm{P}} \xrightarrow{k_6} \mathrm{Hpt} \qquad \mathrm{HK}_{\mathrm{PP}} + \mathrm{Hpt} \xrightarrow{k_5} \mathrm{HK}_{\mathrm{P}0} + \mathrm{Hpt}_{\mathrm{P}}.$$

Note that for notational convenience, we have renamed the relevant rate constants as $k_1, \ldots, k_6$ and denote the concentration of the species as follows:

$$x_1 := [\mathrm{HK}_{00}] \qquad x_2 := [\mathrm{HK}_{\mathrm{P}0}] \qquad x_3 := [\mathrm{HK}_{0\mathrm{P}}] \qquad x_4 := [\mathrm{HK}_{\mathrm{PP}}]$$
$$x_5 := [\mathrm{Hpt}] \qquad x_6 := [\mathrm{Hpt}_{\mathrm{P}}].$$

Under the law of mass-action, we model the dynamics of the concentrations over time by the following system of ordinary differential equations:

$$\dot{x}_1 = k_4 x_3 x_5 - k_1 x_1 \tag{126}$$

$$\dot{x}_2 = k_5 x_4 x_5 + k_1 x_1 - k_2 x_2 \tag{127}$$

$$\dot{x}_3 = k_2 x_2 - k_3 x_3 - k_4 x_3 x_5 \tag{128}$$

$$\dot{x}_4 = k_3 x_3 - k_5 x_4 x_5 \tag{129}$$

$$\dot{x}_5 = k_6 x_6 - k_4 x_3 x_5 - k_5 x_4 x_5 \tag{130}$$

$$\dot{x}_6 = -k_6 x_6 + k_4 x_3 x_5 + k_5 x_4 x_5, \tag{131}$$

Observe that

$$\dot{x}_1 + \dot{x}_2 + \dot{x}_3 + \dot{x}_4 = 0 \quad \text{and} \quad \dot{x}_5 + \dot{x}_6 = 0.$$

It follows that the sums of concentrations $x_1 + x_2 + x_3 + x_4$ and $x_5 + x_6$ are constant over time. This leads to the extra equations

$$x_1 + x_2 + x_3 + x_4 = H, \qquad x_5 + x_6 = T. \tag{132}$$

for some positive total amounts $H, T > 0$.

3.2.2.1 Positive steady states

The positive steady states of the system are the solutions to the equations $\dot{x}_1, \ldots, \dot{x}_6 = 0$, constrained by the conservation laws in eq. (132). Due to the conservation laws, the equation $\dot{x}_6 = 0$ is fulfilled provided $\dot{x}_5 = 0$ is fulfilled, and similarly, $\dot{x}_4 = 0$ is fulfilled provided $\dot{x}_1 = 0, \dot{x}_2 = 0$, and $\dot{x}_3 = 0$ are fulfilled. Therefore, the equations $\dot{x}_6 = 0$ and $\dot{x}_4 = 0$ can be disregarded. Consider first the system of equations given by $\dot{x}_1 = 0, \dot{x}_2 = 0, \dot{x}_3 = 0$, and the first equation in (132). That is, consider the system of equations:

$$0 = k_4 x_3 x_5 - k_1 x_1 \tag{133}$$

$$0 = k_5 x_4 x_5 + k_1 x_1 - k_2 x_2 \tag{134}$$

$$0 = k_2 x_2 - k_3 x_3 - k_4 x_3 x_5 \tag{135}$$

$$H = x_1 + x_2 + x_3 + x_4. \tag{136}$$

This system is linear in x_1, x_2, x_3, x_4 with coefficients involving the rate constants and x_5. We solve it and obtain the following algebraic expressions for x_1, x_2, x_3, x_4 at steady state, depending on the value of x_5 at steady state:

$$x_1 = \frac{k_2 k_4 k_5 H x_5^2}{(k_1 + k_2 k_4) k_5 x_5^2 + k_1 (k_2 + k_3) k_5 x_5 + k_1 k_2 k_3} \tag{137}$$

$$x_2 = \frac{k_1 (k_4 x_5 + k_3) k_5 H x_5}{(k_1 + k_2 k_4) k_5 x_5^2 + k_1 (k_2 + k_3) k_5 x_5 + k_1 k_2 k_3} \tag{138}$$

$$x_3 = \frac{k_1 k_2 k_5 H x_5}{(k_1 + k_2 k_4) k_5 x_5^2 + k_1 (k_2 + k_3) k_5 x_5 + k_1 k_2 k_3} \tag{139}$$

$$x_4 = \frac{k_1 k_2 k_3 H}{(k_1 + k_2 k_4) k_5 x_5^2 + k_1 (k_2 + k_3) k_5 x_5 + k_1 k_2 k_3}. \tag{140}$$

These expressions are positive provided x_5 is positive. We use the second equation in eq. (132) to determine the value at steady state of x_6. Clearly, we have that $x_6 = T - x_5$, which is positive provided $x_5 < T$.

3.2.2.2 The steady state polynomial.

All concentrations are expressed as functions of x_5, and we have not used the equation $\dot{x}_5 = 0$. We can replace the equation $\dot{x}_5 = 0$ by any linear combination of the steady state equations that involves this one. By doing so, the solutions to the equations do not change. We replace it by the equation $\dot{x}_5 + \dot{x}_1 - \dot{x}_4 = 0$. This cancels out the quadratic terms in the equation $\dot{x}_5 = 0$, and we obtain the equation

$$0 = k_6 x_6 - k_1 x_1 - k_3 x_3. \tag{141}$$

Substituting into (141) the values of x_1 and x_3 in (137), (139), and further letting $x_6 = T - x_5$, we obtain that, at steady state, it holds

$$0 = k_6(T - x_5) - \frac{k_1k_2k_4k_5Hx_5^2 + k_1k_2k_3k_5Hx_5}{(k_1 + k_2k_4)k_5x_5^2 + k_1(k_2 + k_3)k_5x_5 + k_1k_2k_3}. \tag{142}$$

By clearing denominators, it can be shown that the positive solutions to (142) agree with the positive solutions to the polynomial

$$\begin{aligned} p(x_5) &= k_6(T - x_5)((k_1 + k_2k_4)k_5x_5^2 + k_1(k_2 + k_3)k_5x_5 + k_1k_2k_3) - k_1k_2k_4k_5Hx_5^2 - k_1k_2k_3k_5x_5 \\ &= (k_1 + k_2)k_4k_5k_6x_5^3 + (k_1(Hk_2k_4 + k_2k_6 + k_3k_6) - T(k_1 + k_2)k_4k_6)k_5x_5^2 \\ &\quad + (k_1k_2k_3(Hk_5 + k_6) - Tk_1(k_2 + k_3)k_5k_6)x_5 - Tk_1k_2k_3k_6 \end{aligned} \tag{143}$$

The polynomial $p(x_5)$ has degree 3. Any root of the polynomial between 0 and T corresponds to a positive steady state. From (142), if $x_5 \geq T$, this results in a positive root of the polynomial, where we would have 0 equal to a negative number, which is a contradiction. Therefore, any positive solution to the polynomial equation must fulfil that $x_5 < T$ and hence provide a positive steady state. The polynomial $p(x_5)$ has at most 3 positive roots. We show in section 3.2.2.4 that there exist choices of rate constants and total amounts such that $p(x_5)$ indeed has 3 positive roots. Therefore, there exist choices of rate constants and total amounts such that the system has 3 positive steady states.

3.2.2.3 Necessary conditions for bistability

Following Descartes' rule of signs, a necessary condition for $p(x_5)$ to have 3 positive roots is that the coefficients of the polynomial have alternating signs. Since the leading coefficient is positive and the independent term is negative, a necessary condition is that the coefficient of degree 2 is negative and the co-efficient of degree 1 is positive, that is:

$$k_1(Hk_2k_4 + k_2k_6 + k_3k_6) < T(k_1 + k_2)k_4k_6, \quad k_1k_2k_3(Hk_5 + k_6) > Tk_1(k_2 + k_3)k_5k_6.$$

To derive the necessary condition for bistability, we consider equation (142) again. We rewrite it as

$$k_6T = k_6x_5 + \frac{k_1k_2k_4k_5Hx_5^2 + k_1k_2k_3k_5Hx_5}{(k_1 + k_2k_4)k_5x_5^2 + k_1(k_2 + k_3)k_5x_5 + k_1k_2k_3}. \tag{144}$$

If the right-hand side of the equation, call it $\varphi(x_5)$, is an increasing function of x_5 for positive x_5, then for any value of T there will be a unique value of x_5 such that

(144) is fulfilled, and hence a unique positive steady state. Since (144) is derived from (141), the function $\varphi(x_5)$ equals $k_6x_5 + k_1x_1 + k_3x_3$, with x_1, x_3 expressed as in (137) and (139). Clearly, k_6x_5 is increasing in x_5. The derivative of $k_1x_1 + k_3x_3$ with respect to x_5 is

$$\frac{Hk_1k_2^2k_5((k_1 - k_3)k_4k_5x_5^2 + 2\,k_1k_3k_4x_5 + k_1k_3^2)}{((k_1 + k_2k_4)k_5x_5^2 + k_1(k_2 + k_3)k_5x_5 + k_1k_2k_3)^2}.$$

The derivative is not necessarily positive for $x_5 > 0$. However, if $k_1 > k_3$ then the derivative of $\varphi(x_5)$ is positive, implying that φ is an increasing function, and, as a consequence, bistability cannot arise for any value of T. To summarise, $k_1 > k_3$ implies that there is no bistability, and therefore, a necessary condition for bistability is that $k_3 > k_1$.

3.2.2.4 Necessary and sufficient conditions for bistability

The conditions given above are only necessary for bistability, but their fulfilment does not guarantee bistability. We provide here necessary and sufficient conditions on all the parameters of the system for bistability. The parameters include the reaction rate constants and the total amounts. To obtain them, we apply Sturm's Theorem:

Theorem 1 (Sturm)**.** *Let $p(x)$ be a real polynomial. Define recursively the Sturm sequence by*

$$p_0(x) = p(x),\ p_1(x) = p'(x), \quad \textit{and} \quad p_{i+1}(x) = -\mathrm{rem}(p_{i-1}, p_i),$$

for $i \geq 1$, where $\mathrm{rem}(p_{i-1}, p_i)$ denotes the reminder of p_{i-1} divided by p_i. The sequence stops when $p_{i+1} = 0$. Let p_m be the last non-zero polynomial.
For $c \in \mathbb{R}$, let $\sigma(c)$ be the number of sign changes in the sequence $p_0(c), \ldots, p_m(c)$. Let $a < b$ and assume that neither a nor b are multiple roots of $p(x)$. Then $\sigma(a) - \sigma(b)$ is the number of distinct roots of $p(x)$ in the interval $(a, b]$.

We are interested in the positive roots of the polynomial $p(x) = p(x_5)$ in (143). In this case, $a = 0$, and we need to take b large enough, which is equivalent to considering instead of the sequence $p_0(b), \ldots, p_m(b)$, the leading coefficients of the polynomials $p_0, \ldots, p_m$. This sequence is written as $p_0(+\infty), \ldots, p_m(+\infty)$. Observe that $a = 0$ is not a root of $p(x)$.

According to the theorem, $\sigma(0) - \sigma(+\infty)$ equals the number of distinct positive roots of $p(x)$. Since $\sigma(0) - \sigma(+\infty) \geq 0$, the number of distinct roots will be 3, that is, we will have three positive steady states, if and only if $\sigma(0) = 3$ and $\sigma(+\infty) = 0$.

We compute the Sturm sequence $p_0(x), \ldots, p_3(x)$ ($p_4(x) = 0$) wherein, for a generic polynomial of degree 3, $p_0(x) = a_0x^3 + a_1x^2 + a_2x + a_3$, the sequence is:

$$\begin{aligned}
p_0(x) &= a_0x^3 + a_1x^2 + a_2x + a_3 \\
p_1(x) &= 3a_0x^2 + 2a_1x + a_2 \\
p_2(x) &= -\frac{6a_0a_2x - 2a_1^2x - 9a_0a_3 - a_1a_2}{9a_0} \\
p_3(x) &= -\frac{9a_0(27a_0^2a_3^2 + 18\,a_0a_1a_2a_3 + 4a_0a_2^3 - 4a_1^3a_3 - a_1^2a_2^2)}{4(3a_0a_2 - a_1^2)^2}.
\end{aligned}$$

In our case, the coefficients are:

$$a_0 = (k_1 + k_2)k_4k_5k_6 > 0$$

$$a_1 = (k_1(Hk_2k_4 + k_2k_6 + k_3k_6) - T(k_1 + k_2)k_4k_6)k_5$$

$$a_2 = (k_1k_2k_3(Hk_5 + k_6) - Tk_1(k_2 + k_3)k_5k_6)$$

$$a_3 = -Tk_1k_2k_3k_6 < 0$$

Hence, $p_0(0) = a_3 < 0$. Therefore, for $\sigma(0) = 3$, we need $p_1(0) > 0, p_2(0) < 0$ and $p_3(0) > 0$. On the other hand,

$$p_0(+\infty) = a_0 > 0 \quad \text{and} \quad p_1(+\infty) = 3a_0 > 0.$$

Therefore, for $\sigma(+\infty) = 0$ we require $p_2(+\infty), p_3(+\infty) > 0$.

The polynomial $p_3(x)$ has degree zero, and hence $p_3(0) = p_3(+\infty)$. Therefore, we are left with 4 conditions on the parameters that fully characterise the region of the parameter space with three steady states, namely $p_1(0), p_3(0), p_2(+\infty) > 0$ and $p_2(0) < 0$. Using that $a_0 > 0$ and $a_3 < 0$, these conditions simplify to the following conditions, where $a_0, \ldots, a_3$ need to be substituted by their expressions in equations **??-??** :

$$\begin{aligned}
a_2 > 0 \qquad & (p_1(0) > 0) \\
9a_0a_3 + a_1a_2 < 0 \qquad & (p_2(0) < 0) \\
27a_0^2a_3^2 + 18a_0a_1a_2a_3 - 4a_0a_2^3 + 4a_1^3a_3 - a_1^2a_2^2 < 0 \qquad & (p_3(0) > 0) \\
-6a_0a_2 + 2a_1^2 > 0 \qquad & (p_2(+\infty) > 0).
\end{aligned}$$

That is, the system has three positive steady states if and only if the 4 inequalities

above are satisfied.

3.2.2.5 Analysis of the steady states of the reaction systems

In summary, solving ODEs representing core hybrid histidine kinase at steady state, shows that positive steady states in the bistable core network are in one-to-one correspondence with the positive roots of the following polynomial (see equation (143)).

$$\begin{aligned} p(Hpt) = k_6(T - Hpt)((k_1 + k_2k_4)k_5Hpt^2 + k_1(k_2 + k_3)k_5Hpt + k_1k_2k_3) - \\ k_1k_2k_4k_5HK_{total}Hpt^2 - k_1k_2k_3k_5Hpt(k_1 + k_2)k_4k_5k_6Hpt^3 + \\ (k_1(HK_{total}k_2k_4 + k_2k_6 + k_3k_6) - Hpt_{total}(k_1 \quad + k_2)k_4k_6)k_5Hpt^2 + \\ (k_1k_2k_3(Hk_5 + k_6) - Hpt_{total}k_1(k_2 + k_3)k_5k_6)Hpt - Hpt_{total}k_1k_2k_3k_6 \end{aligned} \quad (145)$$

The reaction rates are as shown in eqs. (118) and (119) and the parameters T and H stand for the total concentrations of Hpt and hybrid HK respectively. Using that $T = Hpt + Hpt_p$, the polynomial is easily transformed into a polynomial which is a function of Hpt_p. This polynomial has at most 3 positive roots and we showed in section 3.2.2.4 that there exist choices of rate constants and total concentrations such that the system has indeed 3 positive steady states, i.e. displays bistability (figure 3.2C). These results obtained for the core system are used to determine how many steady states the full systems can have. We generalise the above analysis to systems comprising n hybrid HKs. In this case we show that the positive steady states are in one-to-one correspondence with the positive roots of a polynomial of degree $2n+1$ in the concentration of Hpt, implying that there can at most be $2n+1$ positive steady states. The coefficients of the polynomial depend on the reaction rate constants and the total concentrations of each HK and Hpt. We subsequently show that there is always a choice of parameters such that the $2n+1$ roots of the polynomial are all positive, thereby giving a choice of parameters such that the system has $2n+1$ positive steady states (see section 3.2.4.3). Our reasoning involves the use of general results on chemical reaction networks modeled with mass action kinetics. Specifically, we use that if a network admits N steady states, then the network obtained after making some reactions reversible [130] or adding intermediate complex formation [95] can also admit at least N steady states. The following section describes this in more detail.

3.2.3 From core to full models

In previous sections we have analysed several systems based on a core sub network of a larger phosphorelay signalling architecture i.e. we disregarded reverse reactions, hydrolysis reactions, and complex formation in the phosphotransfer reactions. We provide here arguments that guarantee that the properties on the number of steady states of the different core models that we have considered extend to the full models.

3.2.3.1 Theoretical results

The results concerning the number of steady states extend to the full models. Two mathematical results, valid for mass-action kinetics, are used for this claim ([95, 130]):

- Assume that a network has N non-degenerate) positive steady states. If complex formation is taken into account, that is, a reaction is split into two by adding an intermediate, then the new extended network also has N (non-degenerate) positive steady states for some choice of rate constants and total amounts [95].
- Assume that a network has N non-degenerate) positive steady states. If reactions are added to the network, in such a way that the conservation laws of the system are preserved, then the new network also has N non-degenerate) positive steady states for some choice of rate constants and total amounts [130].

The non-degeneracy condition means that the Jacobian is non-singular relative to the stoichiometric compatibility class described by the conservation laws (section 3.2.4.4). This requirement is fulfilled in our case.

3.2.3.2 Full hybrid HK

The full model of a hybrid HK with reversible reactions and hydrolysis reactions, consists of the reactions:

$$\begin{array}{lll}
\mathrm{HK_{00}} \xrightarrow{k_1} \mathrm{HK_{P0}} \xrightarrow{k_2} \mathrm{HK_{0P}} \xrightarrow{k_3} \mathrm{HK_{PP}} & & \mathrm{HK_{0P}} + \mathrm{Hpt} \underset{k_7}{\overset{k_4}{\rightleftharpoons}} \mathrm{HK_{00}} + \mathrm{Hpt_P} \\
\mathrm{HK_{0P}} \xrightarrow{k_9} \mathrm{HK_{00}} & \mathrm{HK_{PP}} \xrightarrow{k_{10}} \mathrm{HK_{P0}} & \mathrm{HK_{PP}} + \mathrm{Hpt} \underset{k_8}{\overset{k_5}{\rightleftharpoons}} \mathrm{HK_{P0}} + \mathrm{Hpt_P} \\
 & \mathrm{RR_P} \xrightarrow{k_{13}} \mathrm{RR} & \mathrm{Hpt_P} + \mathrm{RR} \underset{k_{11}}{\overset{k_{11}}{\rightleftharpoons}} \mathrm{Hpt} + \mathrm{RR_P}.
\end{array} \tag{146}$$

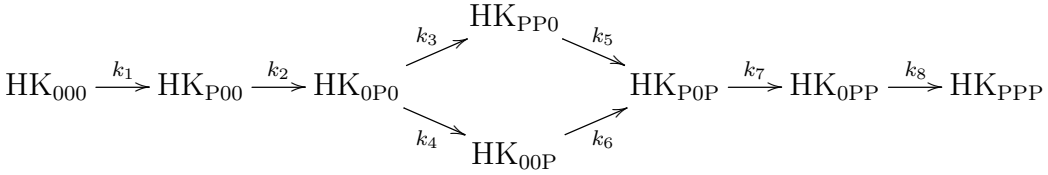
One might also consider complex formation, that is, substitute each phosphotransfer reaction of the form $A + B \rightleftharpoons C + D$ by

$$A + B \rightleftharpoons Y \rightleftharpoons C + D.$$

From section 3.2.2.4, we see that the core phosphorelay with a hybrid HK can have 3 positive steady states. Adding reversibility to some reactions does not change the conservation laws, nor does including hydrolysis reactions. Therefore, by [130], the full phosphorelay model with a hybrid HK can have 3 positive steady states for some choice of rate constants and total amounts. This holds true even if hydrolysis reactions are added on the other phosphorylation sites (that is, the first domain of the hybrid HK and Hpt). Adding complex formation also maintains the maximal number of steady states[95]. In both cases, however, a higher number of steady states might be achievable. The same argument holds for all the models considered in previous sections.

3.2.3.3 Unorthodox HK

Consider now the core model with unorthodox HK, obtained by removing some hydrolysis reactions and reversibility. HK has now three phosphorylatable domains. The reactions of the system are as follows:



$$HK_{00P} + RR \xrightarrow{k_9} HK_{000} + RR_P \qquad HK_{P0P} + RR \xrightarrow{k_{10}} HK_{P00} + RR_P$$

$$HK_{0PP} + RR \xrightarrow{k_{11}} HK_{0P0} + RR_P \qquad HK_{PPP} + RR \xrightarrow{k_{12}} HK_{PP0} + RR_P$$

$$RR_P \xrightarrow{k_{13}} RR$$

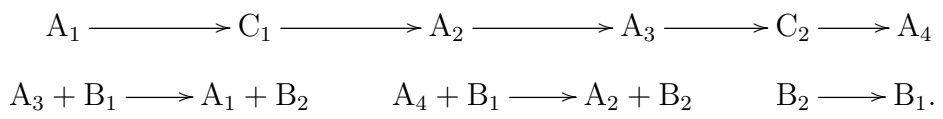
We show here that this model is essentially obtained by adding iteratively species and reactions to the core hybrid HK model in subsection 3.2.2, in the sense allowed by [130] and [95]. Not to confuse notation, we write the model in subsection 3.2.2 with the following labels for each species:

$$A_1 \longrightarrow A_2 \longrightarrow A_3 \longrightarrow A_4$$

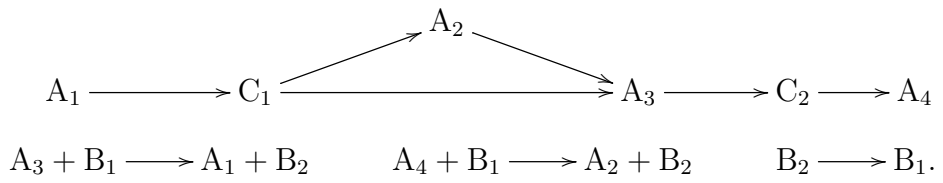
$$A_3 + B_1 \longrightarrow A_1 + B_2 \qquad A_4 + B_1 \longrightarrow A_2 + B_2 \qquad B_2 \longrightarrow B_1.$$

We perform the following modifications to the model:

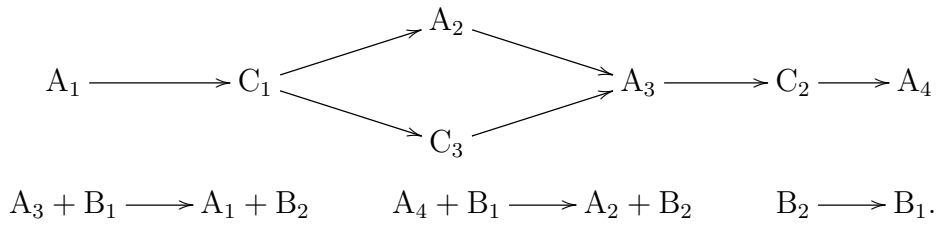
(i) Add species C_1, C_2:



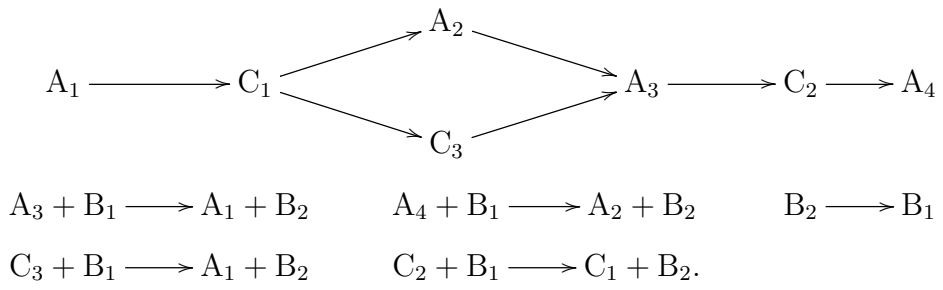
(ii) Add a reaction from C_1 to A_3:



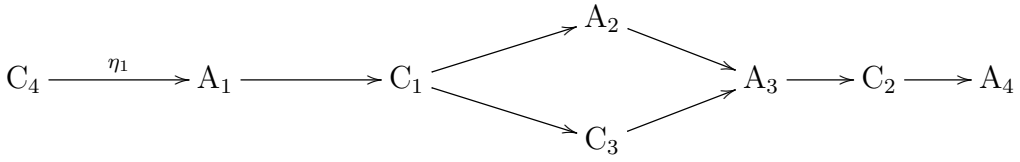
(iii) Add a species between C_1 and A_3:



(iv) Add two reactions:



From [130] or [95], we see that each step preserves the number of positive steady states of the network from the previous step. The network in (iv) is not exactly the core model of an unorthodox HK, but almost. The core model of the unorthodox HK requires an extra species C_5, a reaction $C_5 \to A_1$, and the reaction $\mathrm{C_3 + B_1 \to A_1 + B_2}$ must be substituted by $\mathrm{C_3 + B_1 \to C_5 + B_2}$. That is, the network we would like to have is



$$\mathrm{A_3 + B_1 \longrightarrow A_1 + B_2} \qquad \mathrm{A_4 + B_1 \longrightarrow A_2 + B_2} \qquad \mathrm{B_2 \longrightarrow B_1}$$

$$\mathrm{C_3 + B_1 \xrightarrow{\eta_2} C_4 + B_2} \qquad \mathrm{C_2 + B_1 \longrightarrow C_1 + B_2}.$$

The new species C_4 is not an intermediate, but behaves exactly like one. Consider the ODE system associated to this network. The steady state equation for $[C_4]$ gives

$$[\mathrm{C_4}] = \frac{\eta_2}{\eta_1}[\mathrm{C_3}][\mathrm{B_1}].$$

Plugging this value back into the ODE system to eliminate $[\mathrm{C_4}]$, we obtain the ODE system associated to the network in item (iv). This is the only characteristic of the intermediates that allows one to prove that the number of steady states is maintained. Therefore, the arguments given in [95] hold also for $\mathrm{C_4}$. We conclude that the core model of the unorthodox HK has at least 3 positive steady states for some choice of rate constants and total amounts, because this is the case for the core model for the hybrid HK. This has also been confirmed using CRNT Toolbox (`https://crnt.osu.edu/LecturesOnReactionNetworks`).

Using approaches discussed in previous sections, we examine networks which contain multidomain HKs and share components (proteins) in the network to form a branched signalling network. The aim is to understand the effect that sharing components has on the overall number of steady states that the combined network has. For this, we examine systems where hybrid HKs share a common Hpt and a shared common RR.

3.2.4 The core model for n hybrid HKs competing for the same Hpt

3.2.4.1 Model description

We study here the core system consisting of n hybrid HKs competing for the same Hpt. We call such a system an *nHK-Hpt system* for short. In this case, there are n hybrid HKs, which we denote by HK^i, for $i = 1, \ldots, n$, and we use subindices $00, \mathrm{P0}, 0\mathrm{P}, \mathrm{PP}$ to denote the phosphorylation state of each of them. The set of

reactions given in the previous subsection are reproduced for the n hybrid HKs. That is, for $i = 1, \ldots, n$, the reactions for the transfer of phosphate groups are as follows:

$$\mathrm{HK}^i_{00} \xrightarrow{k_{i,1}} \mathrm{HK}^i_{\mathrm{P}0} \xrightarrow{k_{i,2}} \mathrm{HK}^i_{0\mathrm{P}} \xrightarrow{k_{i,3}} \mathrm{HK}^i_{\mathrm{PP}} \qquad \mathrm{HK}^i_{0\mathrm{P}} + \mathrm{Hpt} \xrightarrow{k_{i,4}} \mathrm{HK}^i_{00} + \mathrm{Hpt}_{\mathrm{P}}$$

$$\mathrm{HK}^i_{\mathrm{PP}} + \mathrm{Hpt} \xrightarrow{k_{i,5}} \mathrm{HK}^i_{\mathrm{P}0} + \mathrm{Hpt}_{\mathrm{P}}$$

and there is further the dephosphorylation reaction

$$\mathrm{Hpt}_{\mathrm{P}} \xrightarrow{k_6} \mathrm{Hpt}.$$

We denote the concentration of the species as follows:

$$x_{i,1} := [\mathrm{HK}^i_{00}], \qquad x_{i,2} := [\mathrm{HK}^i_{\mathrm{P}0}], \qquad x_{i,3} := [\mathrm{HK}^i_{0\mathrm{P}}], \qquad x_{i,4} := [\mathrm{HK}^i_{\mathrm{PP}}],$$
$$x_5 := [\mathrm{Hpt}], \qquad x_6 := [\mathrm{Hpt}_{\mathrm{P}}],$$

for $i = 1, \ldots, n$. Under the law of mass-action, we model the dynamics of the concentrations over time by the following system of ordinary differential equations:

$$\dot{x}_{i,1} = k_{i,4} x_{i,3} x_5 - k_{i,1} x_{i,1} \tag{147}$$
$$\dot{x}_{i,2} = k_{i,5} x_{i,4} x_5 + k_{i,1} x_{i,1} - k_{i,2} x_{i,2} \tag{148}$$
$$\dot{x}_{i,3} = k_{i,2} x_{i,2} - k_{i,3} x_{i,3} - k_{i,4} x_{i,3} x_5 \tag{149}$$
$$\dot{x}_{i,4} = k_{i,3} x_{i,3} - k_{i,5} x_{i,4} x_5 \tag{150}$$
$$\dot{x}_5 = k_6 x_6 - \sum_{j=1}^{n} (k_{j,4} x_{j,3} x_5 + k_{j,5} x_{j,4} x_5) \tag{151}$$
$$\dot{x}_6 = -k_6 x_6 + \sum_{j=1}^{n} (k_{j,4} x_{j,3} x_5 + k_{j,5} x_{j,4} x_5), \tag{152}$$

for $i = 1, \ldots, n$. The system has $n + 1$ conservation laws for $i = 1, \ldots, n$, we have

$$x_{i,1} + x_{i,2} + x_{i,3} + x_{i,4} = H_i \tag{153}$$

for some $H_i > 0$, and for $T > 0$,

$$x_5 + x_6 = T \tag{154}$$

3.2.4.2 Positive steady states

The positive steady states of the system are the solutions to the equations $\dot{x}_{i,1} = 0, \dot{x}_{i,2} = 0, \dot{x}_{i,3} = 0, \dot{x}_{i,4} = 0$, for $i = 1, \ldots, n$, together with $\dot{x}_5 = 0, \dot{x}_6 = 0$, constrained by the conservation laws (153) and (154). We reason as discussed in section 3.2.2.1 to disregard the steady state equations $\dot{x}_6 = 0$ and $\dot{x}_{i,4} = 0$, for $i = 1, \ldots, n$. Using the equations $\dot{x}_{i,1} = 0, \dot{x}_{i,2} = 0, \dot{x}_{i,3} = 0$ and (153), namely

$$\begin{aligned} 0 &= k_{i,4}x_{i,3}x_5 - k_{i,1}x_{i,1} \\ 0 &= k_{i,5}x_{i,4}x_5 + k_{i,1}x_{i,1} - k_{i,2}x_{i,2} \\ 0 &= k_{i,2}x_{i,2} - k_{i,3}x_{i,3} - k_{i,4}x_{i,3}x_5 \\ H_i &= x_{i,1} + x_{i,2} + x_{i,3} + x_{i,4}, \end{aligned} \tag{155}$$

we obtain the algebraic expressions for $x_{i,1}, x_{i,2}, x_{i,3}, x_{i,4}$ at steady state, depending on the value of x_5 at steady state, analogous to the expressions (137)-(140):

$$x_{i,1} = \frac{k_{i,2}k_{i,4}k_{i,5}H_i x_5^2}{(k_{i,1} + k_{i,2}k_{i,4})k_{i,5}x_5^2 + k_{i,1}(k_{i,2} + k_{i,3})k_{i,5}x_5 + k_{i,1}k_{i,2}k_{i,3}} \tag{156}$$

$$x_{i,2} = \frac{k_{i,1}(k_{i,4}x_5 + k_{i,3})k_{i,5}H_i x_5}{(k_{i,1} + k_{i,2}k_{i,4})k_{i,5}x_5^2 + k_{i,1}(k_{i,2} + k_{i,3})k_{i,5}x_5 + k_{i,1}k_{i,2}k_{i,3}} \tag{157}$$

$$x_{i,3} = \frac{k_{i,1}k_{i,2}k_{i,5}H_i x_5}{(k_{i,1} + k_{i,2}k_{i,4})k_{i,5}x_5^2 + k_{i,1}(k_{i,2} + k_{i,3})k_{i,5}x_5 + k_{i,1}k_{i,2}k_{i,3}} \tag{158}$$

$$x_{i,4} = \frac{k_{i,1}k_{i,2}k_{i,3}H_i}{(k_{i,1} + k_{i,2}k_{i,4})k_{i,5}x_5^2 + k_{i,1}(k_{i,2} + k_{i,3})k_{i,5}x_5 + k_{i,1}k_{i,2}k_{i,3}}. \tag{159}$$

These expressions are positive provided x_5 is positive. From (154) we have that $x_6 = T - x_5$, which is positive provided $x_5 < T$.
We replace the steady state equation $\dot{x}_5 = 0$ by $\dot{x}_5 + \sum_{i=1}^{n}(\dot{x}_{i,1} - \dot{x}_{i,4}) = 0$, and we obtain the equation

$$0 = k_6 x_6 - \sum_{i=1}^{n}(k_{i,1}x_{i,1} + k_{i,3}x_{i,3}). \tag{160}$$

Substituting into (160) the values of $x_{i,1}$ and $x_{i,3}$ in (156), (158), and further letting $x_6 = T - x_5$, we obtain that, at steady state

$$0 = k_6(T - x_5) - \sum_{i=1}^{n} \frac{k_{i,1}k_{i,2}k_{i,4}k_{i,5}H_i x_5^2 + k_{i,1}k_{i,2}k_{i,3}k_{i,5}H_i x_5}{(k_{i,1} + k_{i,2}k_{i,4})k_{i,5}x_5^2 + k_{i,1}(k_{i,2} + k_{i,3})k_{i,5}x_5 + k_{i,1}k_{i,2}k_{i,3}}. \tag{161}$$

By clearing denominators, that is, by multiplying equation (161) by

$$\prod_{i=1}^{n}(k_{i,1}+k_{i,2}k_{i,4})k_{i,5}x_5^2+k_{i,1}(k_{i,2}+k_{i,3})k_{i,5}x_5+k_{i,1}k_{i,2}k_{i,3},$$

we obtain a polynomial of degree $2n+1$ in x_5. Any zero of the polynomial, that lies between 0 and T, corresponds to a positive steady state. We argue again that, if $x_5 \geq T$ were a positive zero of the polynomial, equation (161) would give a contradiction. Therefore, any positive solution to the polynomial equation must fulfill that $x_5 < T$ and hence provide a positive steady state.

3.2.4.3 Existence of $2n+1$ positive steady states.

We have shown that the positive steady states of the n HK-Hpt system are determined by the positive solutions to (161). Solving for the positive solutions to this equation is equivalent to solving for the positive solutions to a polynomial of degree $2n+1$. By the fundamental theorem of algebra, a polynomial of degree $2n+1$ has $2n+1$ roots counted with multiplicity. Therefore, such a polynomial can *at most* have $2n+1$ distinct positive real roots.

We show in this section that there exist choices of rate constants k_* and total amounts H_i, T such that the polynomial has exactly $2n+1$ distinct positive real roots. As a consequence, this proves that the nHK-Hpt system admits $2n+1$ positive steady states for some choice of rate constants and total amounts.

The proof consists of a series of simplifications and constructions. A key ingredient of the proof is the following theorem:

Theorem 2 (Kurtz [131]). *Let $m \geq 1$ and let $p(x) = x^{2m+1} - c_1x^{2m} + c_2x^{2m-1} + \cdots + c_{2m}x - c_{2m+1}$ be a polynomial of odd degree $2m+1$ and with $c_i \geq 0$ for all i. Let $c_0 = 1$. If*

$$c_i^2 - 4c_{i-1}c_{i+1} > 0 \tag{162}$$

for all $i = 1, \ldots, 2m$, then $p(x)$ has $2m+1$ distinct positive real roots.

For clarity, we provide the main arguments of our proof in the form of lemmas in Appendix E. First of all observe that the steady states of the system are invariant by multiplication of all rate constants by some scalar $\lambda > 0$. Therefore, we assume that $k_6 = 1$. For simplicity we write x for x_5. We let

$$\alpha_{i,1} = k_{i,1}k_{i,2}k_{i,4}k_{i,5}H_i \tag{163}$$

$$\alpha_{i,2} = k_{i,1}k_{i,2}k_{i,3}k_{i,5}H_i \tag{164}$$

$$\alpha_{i,3} = (k_{i,1} + k_{i,2})k_{i,4}k_{i,5} \tag{165}$$

$$\alpha_{i,4} = k_{i,1}(k_{i,2} + k_{i,3})k_{i,5} \tag{166}$$

$$\alpha_{i,5} = k_{i,1}k_{i,2}k_{i,3}, \tag{167}$$

such that we write

$$\frac{k_{i,1}k_{i,2}k_{i,4}k_{i,5}H_i x^2 + k_{i,1}k_{i,2}k_{i,3}k_{i,5}H_i x}{(k_{i,1} + k_{i,2}k_{i,4})k_{i,5}x_5^2 + k_{i,1}(k_{i,2} + k_{i,3})k_{i,5}x_5 + k_{i,1}k_{i,2}k_{i,3}} = \frac{\alpha_{i,1}x^2 + \alpha_{i,2}x}{\alpha_{i,3}x^2 + \alpha_{i,4}x + \alpha_{i,5}}.$$

Lemma 1. *For any positive values $\alpha_{i,1}, \ldots, \alpha_{i,5} > 0$, there exist $k_{i,1}, \ldots, k_{i,5} > 0$ and $H_i > 0$ such that* (163)-(167) *are fulfilled.*

As a consequence of Lemma 1, there exist rate constants and total amounts such that (161) holds if we can find $\alpha_{i,1}, \ldots, \alpha_{i,5} > 0$ such that

$$0 = -T + x + \sum_{i=1}^{n} \frac{\alpha_{i,1}x^2 + \alpha_{i,2}x}{\alpha_{i,3}x^2 + \alpha_{i,4}x + \alpha_{i,5}}. \tag{168}$$

With this notation, we want to determine the positive real roots of the polynomial obtained by clearing denominators in (168):

$$q(x) = (x-T)\prod_{i=1}^{n}(\alpha_{i,3}x^2+\alpha_{i,4}x+\alpha_{i,5})+\sum_{i=1}^{n}\Big((\alpha_{i,1}x^2+\alpha_{i,2}x)\prod_{j\neq i}(\alpha_{j,3}x^2+\alpha_{j,4}x+\alpha_{j,5})\Big). \tag{169}$$

The coefficient of degree $2n+1$ of $q(x)$ is $\prod_{i=1}^{n}\alpha_{i,3}$ and the independent term of $q(x)$ is $-T\prod_{i=1}^{n}\alpha_{i,5}$. We would like to apply Theorem 2 to such a polynomial. To this end, the coefficients of monomials with even degree should be negative and those of odd degree should be positive. The latter is guaranteed if $\alpha_{i,4} = 0$, while the former if $\alpha_{i,1} = 0$. Setting these two constants to zero, for $i = 1, \ldots, n$, does not change the degree of the polynomial.

By the continuity of the isolated roots of a polynomial as functions of the coefficients of the polynomial, if we can find $\alpha_{i,2}, \alpha_{i,3}, \alpha_{i,5} > 0$ such that with $\alpha_{i,4} = 0, \alpha_{i,1} = 0$, the polynomial $q(x)$ has $2n+1$ distinct positive real roots, then for $\alpha_{i,4}, \alpha_{i,1}$ small enough, the polynomial $q(x)$ still has $2n+1$ distinct positive real roots.

This is what we do next. We set $\alpha_{i,4} = 0, \alpha_{i,1} = 0$ and further $\alpha_{i,3} = 1$ for all $i = 1, \ldots, n$, and $T = 1$. To ease the notation, we write $a_i = \alpha_{i,2}$ and $b_i = \alpha_{i,5}$, such that the polynomial of interest is

$$p(x) = (x-1)\prod_{i=1}^{n}(x^2 + b_i) + \sum_{i=1}^{n}\Big(a_i x \prod_{j\neq i}(x^2 + b_j)\Big). \tag{170}$$

We denote by $[n] = \{1, \ldots, n\}$. In the next lemma we describe the coefficients of $p(x)$. The form of the coefficients depends on the parity of the degree of the coefficient. Therefore the coefficients take two different forms, one for even subindices, that is $i = 2k$, and one for odd subindices, that is $i = 2k + 1$.

Lemma 2. *Let the polynomial* (170) *be written as* $p(x) = c_0 x^{2n+1} + c_1 x^{2n} + c_2 x^{2n-1} + \cdots + c_{2n}x + c_{2n+1}$. *Then it holds that:*

$$c_{2k+1} = -\sum_{\{j_1,\ldots,j_k\}\subseteq[n]}\prod_{\ell=1}^{k} b_{j_\ell}$$

$$c_{2k} = \sum_{\{j_1,\ldots,j_k\}\subseteq[n]}\prod_{\ell=1}^{k} b_{j_\ell} + \sum_{i=1}^{n} a_i \sum_{\{j_1,\ldots,j_{k-1}\}\subseteq[n]\setminus\{i\}}\prod_{\ell=1}^{k-1} b_{j_\ell}$$

for $k = 0, \ldots, n$, *with the convention that the sum and the product over the empty set equals* 1.

For example, for $n = 1$ we have

$$p(x) = x^3 - x^2 + (a_1 + b_1)x - b_1,$$

while for $n = 2$ we have

$$p(x) = x^5 - x^4 + (a_1 + a_2 + b_1 + b_2)x^3 - (b_1 + b_2)x^2 + (a_1 b_2 + a_2 b_1 + b_1 b_2)x - b_1 b_2.$$

All that is left is to show that we can find b_i, a_i such that the polynomial $p(x)$ satisfies the inequalities in Theorem 2. This is the content of the following lemmas. We provide in Lemma 3 a choice of constants b_i such that the inequalities (162) are fulfilled for even indices i, that is, $i = 2k$ for some k. In Lemma 4 we provide a choice of constants a_i such that the inequalities (162) are fulfilled for odd indices i, that is, $i = 2k + 1$ for some k.

Lemma 3. *Fix arbitrary* $a_1, \ldots, a_n > 0$ *and define* $b_i = \frac{a_i^2}{4}$ *for* $i = 1, \ldots, n$. *Then,*

$$c_{2k}^2 - 4c_{2k-1}c_{2k+1} > 0$$

for all $k = 1, \ldots, n$.

Lemma 4. *Let $M > 0$ and $a_n > 0$. For $i = 1, \ldots, n$, define $a_i = \frac{a_n}{M^{i-1}}$ and $b_i = \frac{a_i^2}{4}$. Then, for M large enough and a_n small enough, it holds that*

$$c_{2k+1}^2 - 4c_{2k}c_{2k+2} > 0$$

for all $k = 0, \ldots, n-1$.

We are ready to prove the main result on the number of positive steady states.

Theorem 3. *For any $n \geq 1$, there exists a choice of rate constants $k_6 > 0$, $k_{i,1}, k_{i,2}, k_{i,3}, k_{i,4} > 0$ and total amounts $T, H_i > 0$, for $i = 1, \ldots, n$, such that the nHK-Hpt system has $2n+1$ distinct positive steady states.*

Proof. Pick $a_n, M > 0$ and define b_i, a_i, for $i = 1, \ldots, n$, as in Lemma 4. By lemmas 4 and 3, by choosing M large enough and a_n small enough, the inequalities (162) hold. We set $\alpha_{i,2} = a_i, \alpha_{i,5} = b_i, \alpha_{i,3} = 1$, for $i = 1, \ldots, n$, and $T = 1$. Then $p(x)$ in (170) has $2n+1$ distinct positive real roots. We choose $\alpha_{i,1}, \alpha_{i,4} > 0$ small enough such that the polynomial $q(x)$ in (169) has $2n+1$ distinct positive real roots. We set $k_6 = 1$. By construction, any choice $k_{i,1}, k_{i,2}, k_{i,3}, k_{i,4} > 0$ and $H_i > 0$ such that (163)-(167) are fulfilled provides a set of parameters with $2n+1$ distinct positive steady states. Such a choice exists by Lemma 1. □

Observe that the proof is constructive. It gives a procedure to find sets of parameters with the maximal number of steady states. The several checks that the proof requires are easily implemented using most available mathematical software to solve equations (e.g. Maple, Mathematica).

In general, we have observed that given any polynomial $u(x)$ with $2n+1$ distinct positive real roots we can find a_i, b_i such that the coefficients of $u(x)$ agree with c_i in Lemma 2, even if $u(x)$ does not fulfil the conditions of Theorem 2. Such a_i, b_i can be found using mathematical software like Maple.

3.2.4.4 n unstable steady states

In the subsection we show that, considering the $2n+1$ steady states ordered increasingly by their value $x = x_5$, then the steady states number $2, 4, \ldots, 2n$ are unstable relative to the stoichiometric compatibility class [132] they belong to, that is, relative to the invariant subspaces described by the conservation laws (153) and (154).

Since the nHK-Hpt system has $4n+2$ variables and $n+1$ conservation laws, the Jacobian of f in $\dot{x} = f(x)$ always has $n+1$ zero eigenvalues. The remaining $3n+1$ eigenvalues (which could include zero) have corresponding eigenvectors

in the stoichiometric subspace and dictate the dynamics around the steady state and within the stoichiometric compatibility class. If the steady state is locally stable relative to the stoichiometric compatibility class, then the product of these $3n+1$ eigenvalues has sign $(-1)^{3n+1}$. Therefore, if the sign of the product of these eigenvalues is $(-1)^{3n}$, then the steady state is necessarily locally unstable relative to the stoichiometric compatibility class. We argue in the proof of the next theorem that this is the case for the steady states in even position $2, 4, \ldots, 2n$.

Theorem 4. *The $2, 4, \ldots, 2n$-th steady states are unstable relative to the stoichiometric compatibility class.*

Proof. We order the variables of the system as $x_{1,1}, x_{1,2}, x_{1,3}, x_{1,4}, \ldots, x_{n,1}, x_{n,2}, x_{n,3}, x_{n,4}, x_6, x_5$. It follows from [133, Prop 5.3] that the product of the $3n+1$ eigenvalues of the Jacobian with eigenvectors in the stoichiometric space agrees with the determinant of the Jacobian of the function $g\colon \mathbb{R}^{4n+2} \to \mathbb{R}^{4n+2}$ where

$$g_{4(i-1)+1}(x) = x_{i,1} + x_{i,2} + x_{i,3} + x_{i,4} - H_i \tag{171}$$
$$g_{4(i-1)+2}(x) = k_{i,5}x_{i,4}x_5 + k_{i,1}x_{i,1} - k_{i,2}x_{i,2} \tag{172}$$
$$g_{4(i-1)+3}(x) = k_{i,2}x_{i,2} - k_{i,3}x_{i,3} - k_{i,4}x_{i,3}x_5 \tag{173}$$
$$g_{4(i-1)+4}(x) = x_5 + k_{i,3}x_{i,3} - k_{i,5}x_{i,4}, \tag{174}$$

for $i = 1, \ldots, n$ and

$$g_{4n+1}(x) = x_5 + x_6 - T,$$
$$g_{4n+2}(x) = k_6x_6 - \sum_{j=1}^{n}(k_{j,4}x_{j,3} + k_{j,5}x_{j,4})x_5.$$

We now apply the method described in [97], to determine the sign of the determinant of the Jacobian of g from iterative eliminations. One can check that the expressions in (156)-(159) are obtained from iteratively eliminating $x_{i,1}, \ldots, x_{i,4}$ from the equations $g_{4(i-1)+1}(x) = \cdots = g_{4(i-1)+4}(x) = 0$, which are equivalent to (155), and which correspond to the conservation law together with $\dot{x}_{i,2} = \dot{x}_{i,3} = \dot{x}_{i,4} = 0$.

Let $q(x_5)$ be the polynomial obtained after clearing denominators in (161). Then, by [97], the sign of the determinant of the Jacobian of g at a steady state agrees with the sign of the derivative of $q(x_5)$, $q'(x_5)$, times $(-1)^{3n}$. Therefore, if $q'(x_5)$ is positive, then the corresponding steady state is locally unstable. Since $q(0)$ is positive, the first real root of $q(x_5)$ has negative derivative, and then the signs alternate. Therefore, the

steady states corresponding to the $2, 4, \ldots, 2n$-th roots are locally unstable relative to the stoichiometric compatibility class. □

3.2.5 Multiple phosphorelays with a common RR

3.2.5.1 The core model for M hybrid phosphorelays competing for an RR

In the previous section we showed that unlimited multistationarity arises by increasing the number of hybrid HKs that compete for the same Hpt. In this section we show that the same statement holds when the competition occurs at the level of the response regulator RR. For example, we show that the system consisting of two core hybrid phosphorelays, complete with their own Hpt, competing for the same RR can have up to 9 steady states. This system consists of two 1HK-Hpt systems, where the two independent Hpt's donate their phosphate group to the same RR. We study such a system with full generality. We allow each of the phosphorelays to have multiple hybrid HKs, as studied in the previous sections.

For example, consider a system with two His-containing phosphotransfer proteins Hpt^1 and Hpt^2 that transfer the phosphate group to the same RR. Assume, for instance, that Hpt^1 receives the phosphate group from two hybrid HKs, $\text{HK}^{1,1}$ and $\text{HK}^{1,2}$, and that Hpt^2 receives the phosphate group from one hybrid HK, $\text{HK}^{2,1}$. The first upper index of $\text{HK}^{i,j}$ indicates the Hpt index, and the second index indicates the index of the HK in the nHK-Hpt subsystem. By using the notation introduced in the previous section to denote phosphorylated sites, the reactions of this example system are as follows:

(i) Reactions within each HK:

$$\text{HK}_{00}^{1,1} \xrightarrow{k_{1,1,1}} \text{HK}_{\text{P}0}^{1,1} \xrightarrow{k_{1,1,2}} \text{HK}_{0\text{P}}^{1,1} \xrightarrow{k_{1,1,3}} \text{HK}_{\text{PP}}^{1,1}$$

$$\text{HK}_{00}^{1,2} \xrightarrow{k_{1,2,1}} \text{HK}_{\text{P}0}^{1,2} \xrightarrow{k_{1,2,2}} \text{HK}_{0\text{P}}^{1,2} \xrightarrow{k_{1,2,3}} \text{HK}_{\text{PP}}^{1,2}$$

$$\text{HK}_{00}^{2,1} \xrightarrow{k_{2,1,1}} \text{HK}_{\text{P}0}^{2,1} \xrightarrow{k_{2,1,2}} \text{HK}_{0\text{P}}^{2,1} \xrightarrow{k_{2,1,3}} \text{HK}_{\text{PP}}^{2,1}$$

(ii) Phosphotransfer from $\text{HK}^{1,1}$ and $\text{HK}^{1,2}$ to Hpt^1, and phosphotransfer from $\text{HK}^{2,1}$ to Hpt^2:

$$\mathrm{HK}^{1,1}_{0\mathrm{P}} + \mathrm{Hpt}^1 \xrightarrow{k_{1,1,4}} \mathrm{HK}^{1,1}_{00} + \mathrm{Hpt}^1_{\mathrm{P}}, \qquad \mathrm{HK}^{1,1}_{\mathrm{PP}} + \mathrm{Hpt}^1 \xrightarrow{k_{1,1,5}} \mathrm{HK}^{1,1}_{\mathrm{P0}} + \mathrm{Hpt}^1_{\mathrm{P}}$$

$$\mathrm{HK}^{1,2}_{0\mathrm{P}} + \mathrm{Hpt}^1 \xrightarrow{k_{1,2,4}} \mathrm{HK}^{1,2}_{00} + \mathrm{Hpt}^1_{\mathrm{P}}, \qquad \mathrm{HK}^{1,2}_{\mathrm{PP}} + \mathrm{Hpt}^1 \xrightarrow{k_{1,2,5}} \mathrm{HK}^{1,2}_{\mathrm{P0}} + \mathrm{Hpt}^1_{\mathrm{P}}$$

$$\mathrm{HK}^{2,1}_{0\mathrm{P}} + \mathrm{Hpt}^2 \xrightarrow{k_{2,1,4}} \mathrm{HK}^{2,1}_{00} + \mathrm{Hpt}^2_{\mathrm{P}}, \qquad \mathrm{HK}^{2,1}_{\mathrm{PP}} + \mathrm{Hpt}^2 \xrightarrow{k_{2,1,5}} \mathrm{HK}^{2,1}_{\mathrm{P0}} + \mathrm{Hpt}^2_{\mathrm{P}}$$

(iii) Phosphotransfer from Hpt^1 and Hpt^2 to RR:

$$\mathrm{Hpt}^1_{\mathrm{P}} + \mathrm{RR} \xrightarrow{k_{1,6}} \mathrm{Hpt}^1 + \mathrm{RR}_{\mathrm{P}} \qquad \mathrm{Hpt}^2_{\mathrm{P}} + \mathrm{RR} \xrightarrow{k_{2,6}} \mathrm{Hpt}^2 + \mathrm{RR}_{\mathrm{P}}$$

(iv) Dephosphorylation reaction for RR:

$$\mathrm{RR}_{\mathrm{P}} \xrightarrow{k_7} \mathrm{RR}.$$

3.2.5.2 Multiple positive steady states

In general, consider M multiple hybrid phosphorelays competing for the same RR. We let n_i be the number of hybrid HKs of the ith system. That is, each phosphorelay consists of a n_iHK-Hpt system with a further phosphotransfer to RR. In the above example, we have $M = 2$ and $n_1 = 2, n_2 = 1$. We let $\mathrm{HK}^{i,j}$ denote the jth hybrid kinase of the ith multiple HK-Hpt system, where j runs from 1 to n_i, and we let Hpt^i denote the Hpt of the system. We denote the reaction rates by $k_{i,j,*}$ for those involving $\mathrm{HK}^{i,j}$, the dephosphorylation reaction for Hpt^i by $k_{i,6}$ and the dephosphorylation reaction for RR by k_7. The system has conserved total amounts of $\mathrm{HK}^{i,j}$, Hpt^i and RR. We denote the concentration of the species as follows:

$$x_{i,j,1} := [\mathrm{HK}^{i,j}_{00}], \quad x_{i,j,2} := [\mathrm{HK}^{i,j}_{\mathrm{P0}}], \quad x_{i,j,3} := [\mathrm{HK}^{i,j}_{0\mathrm{P}}], \quad x_{i,j,4} := [\mathrm{HK}^{i,j}_{\mathrm{PP}}],$$
$$x_{i,5} := [\mathrm{Hpt}^i], \quad x_{i,6} := [\mathrm{Hpt}^i_{\mathrm{P}}], \quad x_7 = [\mathrm{RR}], \quad x_8 = [\mathrm{RR}_{\mathrm{P}}].$$

for $i = 1, \ldots, M$, $j = 1, \ldots, n_i$. Under the law of mass-action, we model the dynamics of the concentrations over time by the following system of ordinary differential equations:

$$\dot{x}_{i,j,1} = k_{i,j,4}x_{i,j,3}x_{i,5} - k_{i,j,1}x_{i,j,1} \tag{175}$$

$$\dot{x}_{i,j,2} = k_{i,j,5}x_{i,j,4}x_{i,5} + k_{i,j,1}x_{i,j,1} - k_{i,j,2}x_{i,j,2} \tag{176}$$

$$\dot{x}_{i,j,3} = k_{i,j,2}x_{i,j,2} - k_{i,j,3}x_{i,j,3} - k_{i,j,4}x_{i,j,3}x_{i,5} \tag{177}$$

$$\dot{x}_{i,j,4} = k_{i,j,3}x_{i,j,3} - k_{i,5}x_{i,j,4}x_{i,5} \tag{178}$$

$$\dot{x}_{i,5} = k_{i,6}x_{i,6}x_7 - \sum_{\ell=1}^{n_i}(k_{i,\ell,4}x_{i,\ell,3} + k_{i,\ell,5}x_{i,\ell,4})x_{i,5} \tag{179}$$

$$\dot{x}_{i,6} = -k_{i,6}x_{i,6}x_7 + \sum_{\ell=1}^{n_i}(k_{i,\ell,4}x_{i,\ell,3} + k_{i,\ell,5}x_{i,\ell,4})x_{i,5}, \tag{180}$$

$$\dot{x}_7 = k_7x_8 - \sum_{\ell=1}^{M} k_{\ell,6}x_{\ell,6}x_7 \tag{181}$$

$$\dot{x}_8 = -k_7x_8 + \sum_{\ell=1}^{M} k_{\ell,6}x_{\ell,6}x_7, \tag{182}$$

for $i = 1, \ldots, M$, $j = 1, \ldots, n_i$. Further, we have the following conservation law equations:

$$x_{i,j,1} + x_{i,j,2} + x_{i,j,3} + x_{i,j,4} = H_{i,j}, \qquad x_{i,5} + x_{i,6} = T_i, \qquad x_7 + x_8 = R.$$

For each i, j, solving for $\dot{x}_{i,j,1} = 0, \dot{x}_{i,j,2} = 0, \dot{x}_{i,j,3} = 0$ and the conservation law with $H_{i,j}$, we get equalities analogous to (156)-(159), where we replace the subindex i by the pair i, j. The steady state concentrations $x_{i,j}$ are expressed in terms of $x_{i,5}$ and are positive provided $x_{i,5}$ is positive.

We further have $x_{i,6} = T_i - x_{i,5}$. The equation $\dot{x}_{i,5} + \sum_{j=1}^{n_i}(\dot{x}_{i,j,1} - \dot{x}_{i,j,4}) = 0$ replaces $\dot{x}_{i,5} = 0$ and leads to the steady state equation

$$0 = k_{i,6}x_7x_{i,6} - \sum_{j=1}^{n_i}(k_{i,j,1}x_{i,j,1} + k_{i,j,3}x_{i,j,3}), \tag{183}$$

which transforms, as in (161), into the equation

$$0 = k_{i,6}x_7(T_i - x_{i,5})$$
$$- \sum_{j=1}^{n_i} \frac{k_{i,j,1}k_{i,j,2}k_{i,j,5}H_{i,j}x_{i,5}(k_{i,j,4}x_{i,5} + k_{i,j,3})}{(k_{i,j,1} + k_{i,j,2})k_{i,j,4}k_{i,j,5}x_{i,5}^2 + k_{i,j,1}(k_{i,j,2} + k_{i,j,3})k_{i,j,5}x_{i,5} + k_{i,j,1}k_{i,j,2}k_{i,j,3}}, \tag{184}$$

for $i = 1, \ldots, M$.

Finally, using the total amount R and $\dot{x}_7 = 0$, we obtain:

$$x_7 = \frac{k_7 R}{k_7 + \sum_{\ell=1}^{M} k_{\ell,6} x_{\ell,6}} = \frac{k_7 R}{k_7 + \sum_{\ell=1}^{M} k_{\ell,6}(T_\ell - x_{\ell,5})}$$
$$x_8 = \frac{R \sum_{\ell=1}^{M} k_{\ell,6} x_{\ell,6}}{k_7 + \sum_{\ell=1}^{M} k_{\ell,6} x_{\ell,6}} = \frac{R \sum_{\ell=1}^{M} k_{\ell,6}(T_\ell - x_{\ell,5})}{k_7 + \sum_{\ell=1}^{M} k_{\ell,6}(T_\ell - x_{\ell,5})}.$$

Observe that both x_7, x_8 are positive provided $x_{i,5} < T_i$ for all i. By substituting the expression for x_7 into (184), we deduce that the steady states of the system are found by finding positive solutions to (184) in $x_{1,5}, \ldots, x_{M,5}$. The value at steady state of the other concentrations are found using the expressions above. Recall that, as seen in the previous section, a positive solution to (184) satisfies $x_{i,5} < T_i$ and hence $x_{i,6}$ is positive.

Theorem 5. *Consider the system with M multiple hybrid phosphorelays competing for the same RR, and let n_i be the number of hybrid HKs of the ith system. Then there exists a choice of rate constants and total amounts such that the system has $\prod_{i=1}^{M} 2n_i + 1$ positive steady states.*

Proof. For each $i = 1, \ldots, M$, fix parameters $k_{i,j,1}, \ldots, k_{i,j,5}, H_{i,j}, T_i$ such that the n_iHK-Hpt system has $2n_i+1$ steady states, when the dephosphorylation rate constant for Hpt^i in the isolated system is set to one. By Theorem 3 such a choice exists. With this choice, let

$$A_i := \sum_{j=1}^{n_i} \frac{k_{i,j,1} k_{i,j,2} k_{i,j,5} H_{i,j} x_{i,5}(k_{i,j,4} x_{i,5} + k_{i,j,3})}{(k_{i,j,1} + k_{i,j,2}) k_{i,j,4} k_{i,j,5} x_{i,5}^2 + k_{i,j,1}(k_{i,j,2} + k_{i,j,3}) k_{i,j,5} x_{i,5} + k_{i,j,1} k_{i,j,2} k_{i,j,3}},$$

which only depends on $x_{i,5}$. Consider the map

$$\begin{aligned} \varphi\colon \mathbb{R} \times \mathbb{R}^M &\to \mathbb{R}^M \\ (S, x_{1,5}, \ldots, x_{M,5}) &\mapsto \left(\frac{(T_i - x_{i,5})}{1 + S \sum_{\ell=1}^{M} (T_\ell - x_{\ell,5})} - A_i \right)_{i=1,\ldots,M}. \end{aligned}$$

For $S = 0$, the ith component of $\varphi(0, x_{1,5}, \ldots, x_{M,5}) = 0$ is $T_i - x_{i,5} - A_i = 0$, which is the steady state equation (161) for the i-th system $\mathrm{HK}^{i,*}$ and Hpt^i with our choice of rate constants. By the above choice, such equation has $2n_i + 1$ positive solutions. Since the ith component of $\varphi(0, x_{1,5}, \ldots, x_{M,5})$ depends only on $x_{i,5}$, it follows that the equation $\varphi(0, x_{1,5}, \ldots, x_{M,5}) = 0$ has $\prod_{i=1}^{M} 2n_i+1$ positive solutions.

The Jacobian of $\varphi(0, x_{1,5}, \ldots, x_{M,5})$ at a solution $(x^*_{1,5}, \ldots, x^*_{M,5})$ is a diagonal matrix, whose ith entry is

$$\frac{\partial}{\partial x_{i,5}}\varphi_i(0, x_{1,5}, \ldots, x_{M,5}) = \frac{\partial}{\partial x_{i,5}}(T_i - x_{i,5} - A_i)$$

evaluated at $x^*_{i,5}$. Since for each i, our set of positive solutions to $T_i - x_{i,5} - A_i = 0$ is maximal in number, there are no multiple solutions and such a derivative is non-zero.

Therefore, the Jacobian of $\varphi(0, x_{1,5}, \ldots, x_{M,5})$ at a solution is non-singular.This allows us to apply the implicit function theorem to ensure that for $S > 0$ small enough, the equation $\varphi(0, x_{1,5}, \ldots, x_{M,5}) = 0$ has precisely $\prod_{i=1}^{M} 2n_i + 1$ positive solutions.

Fix any such $S > 0$ and define $k_{i,6} = S$, $R = \frac{1}{S}$, and $k_7 = 1$. Then

$$\frac{(T_i - x_{i,5})}{1 + S\sum_{\ell=1}^{M}(T_\ell - x_{\ell,5})} = \frac{k_7 k_{i,6} R(T_i - x_{i,5})}{k_7 + \sum_{\ell=1}^{M} k_{\ell,6}(T_\ell - x_{\ell,5})},$$

which corresponds to the steady state equation (184). With this choice, the system has $\prod_{i=1}^{M} 2n_i + 1$ positive steady states.

□

The proof gives a constructive way to find sets of parameters with $\prod_{i=1}^{M} 2n_i + 1$ positive steady states. We fix parameters for the individual nHK-Hpt systems that have $2n_i + 1$ steady states, when the dephosphorylation rate constant for Hpt^i in the isolated system is set to one. Then, let $k_7 = 1$, $k_{i,6} = 1/R$ and increase R until the system has the desired number of steady states.

3.3 Summary of Findings

To understand the effect of multidomain sensor histidine kinases on the dynamics of phosphorelays and test the question if multidomain HKs embed multistability, we created mathematical models of a phosphorelays containing regular, hybrid and unorthodox HKs (see section 3.2). All of these models implemented reactions which have been shown to occur in experimental studies [82, 107] and included in previous theoretical models of phosphorelays [3, 37, 85]. Analysing the resulting chemical reaction systems for all the three models using CRNT toolbox revealed that a network with a regular HK does not fulfil the theoretical requirements for bistability, while systems with hybrid and unorthodox HKs do (see section 3.2).

To better understand the source of this bistability in multidomain HKs, we focused on a hybrid HK (represented as an entity with four states HK_{OO}, HK_{PO}, HK_{OP} and HK_{PP}) containing phosphorelay (figure 3.2). By identifying a minimal (bistable) sub network of the overall phosphorelay architecture involving this hybrid HK (figure 3.2 (a)) and analytically solving the steady-state equations arising from the set of ordinary differential equations describing the dynamics of this core system, we derive the set of necessary and sufficient conditions as a function of reaction rates and total protein concentrations in the network (see sections 3.2.2.3 & 3.2.2.4).

3.3.1 Reactions among histidine kinase domains and their downstream target gives rise to interconnected feedback loops

While the mathematical complexity of these conditions does not permit a simple biological interpretation, one notable necessary condition for bistability in this core model is that the rate constant of autophosphorylation of the hybrid-HK must be higher when the REC site is phosphorylated compared to when it is not (i.e. the auto-phosphorylation of the OP state must be higher than that of the OO state, $k_3 > k_1$). Interestingly, these two reactions drive two connected feedback loops, where one loop cycles from the OO state, to PO, OP, and then back to OO, while the other cycles from the PO state, to OP, PP, and then back to PO (figure 3.2). This observation allows an intuitive understanding of bistability in this core system. At low signal and high Hpt levels, the auto-phosphorylation of the HK can be balanced between a flow of phosphate groups through Hpt, allowing the first loop I (see figure 3.2) to dominate the dynamics. As the signal increases and Hpt is consumed more and more, this balance is increasingly disrupted and there is suddenly not enough Hpt to absorb all of the phosphate groups from the OP state. This then allows OP to increasingly undergo auto-phosphorylation, which happens faster under the condition of $k_3 > k_1$, and leads to the second loop (PO-OP-PP-PO) to start dominating. This results in a sudden rise in PP and the phosphotransfer rate to Hpt, overwhelming the latter and causing its phosphorylated state to make a sudden jump. This jump is the bifurcation point that we observe in the system dynamics. When loop I dominates, the system is at the lower steady state and when loop II dominates, the system occupies the higher steady state. We find that this intuitive narrative fits with the observed temporal and steady state concentrations of the different phosphorylation states of the HK (figure 3.2) and also explains the effect of increasing the ratio k_3/k_1 on the system dynamics (figure 3.3).

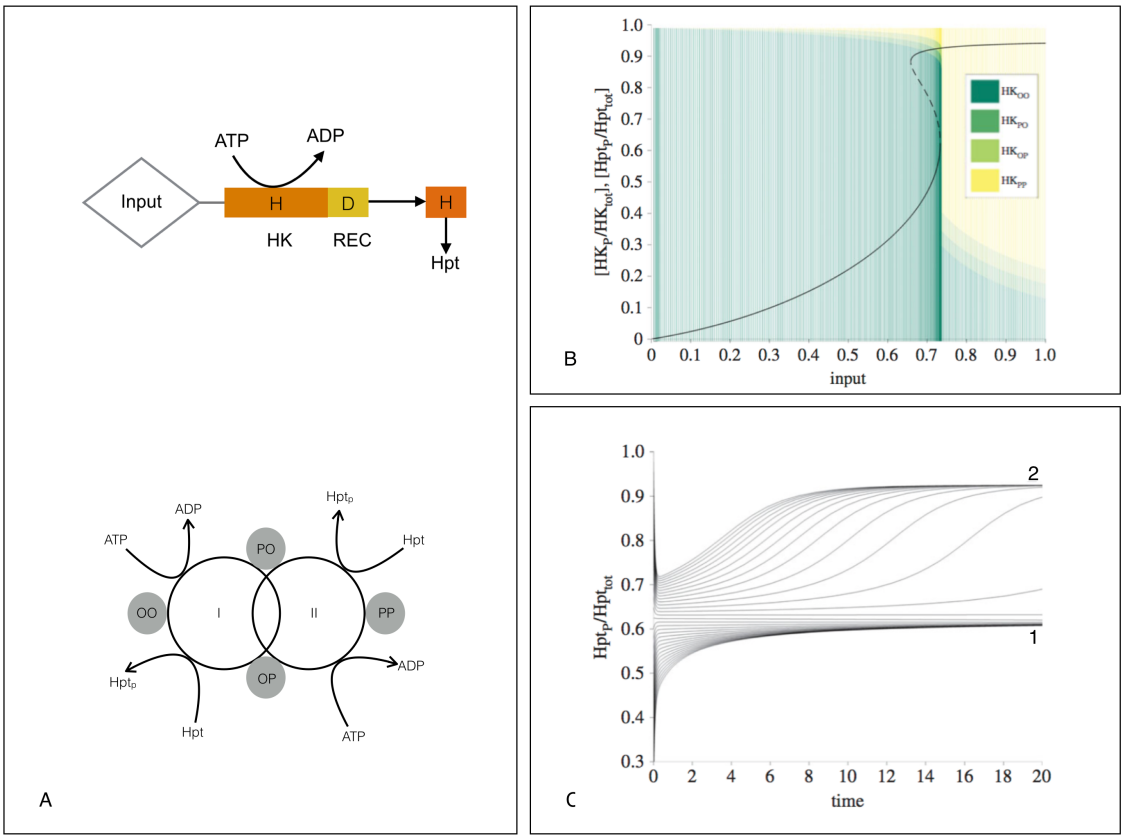


Figure 3.2: Schematic representation showing the minimal core reaction network utilising a hybrid HK that maintains the capacity for multistability. It contains two connected feedback loops which are responsible for the bistability in the core network (A). Plot (B) shows the fraction of steady state phosphorylated Hpt levels for a given input value (appendix F). The change in input level is simulated by varying the auto-phosphorylation rate constants of HK, k_1 and k_3, while keeping the ratio k_3/k_1 fixed. Solid and dotted lines indicate stable and unstable steady states respectively. The steady state fraction of different phosphoforms of HK are shown as a stacked bar (overlay). Here, HK_{PP} is populated only above a threshold input level. Time series plot showing the phosphorylated Hpt levels approaching two different stable steady states depending on the initial conditions (C).

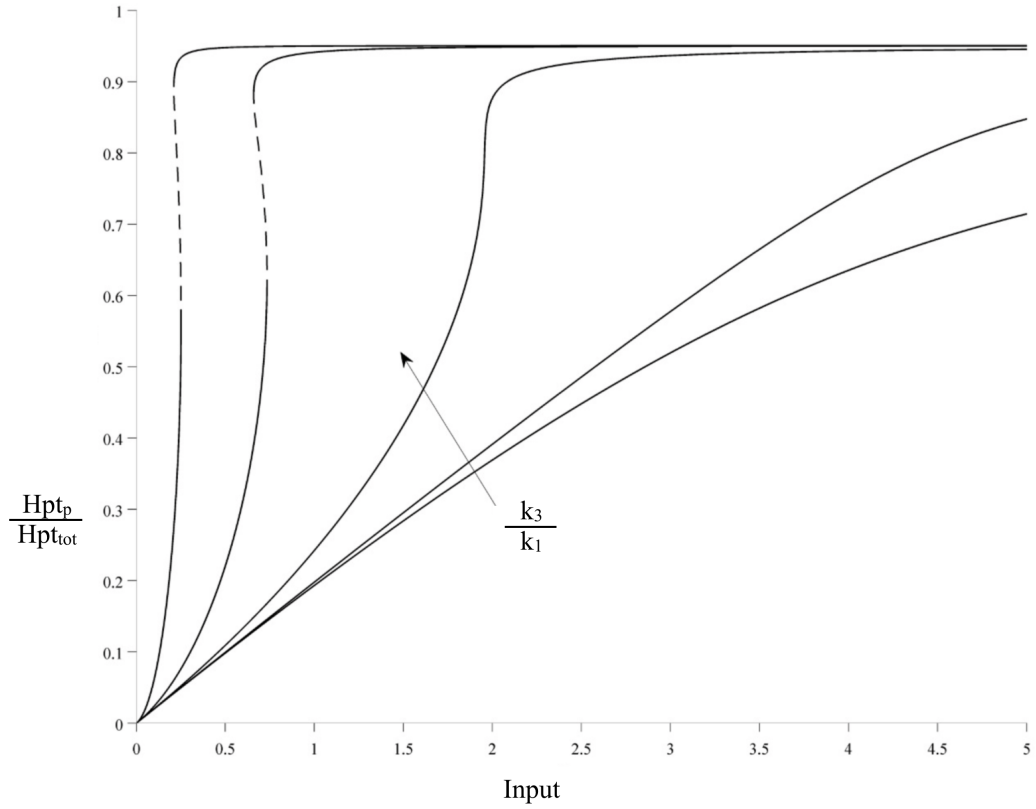


Figure 3.3: Effect of varying the ratio k_3/k_1 on the signal-response plot for the minimal core system shown in Figure 3.2. The signal-response plot shows the fraction of phosphorylated Hpt at steady state while varying the auto-phosphorylation rate constants of HK, k_1 and k_3, such that the ratio k_3/k_1 is fixed for each signal-response plot. The solid and dotted lines indicate stable and unstable steady states respectively.

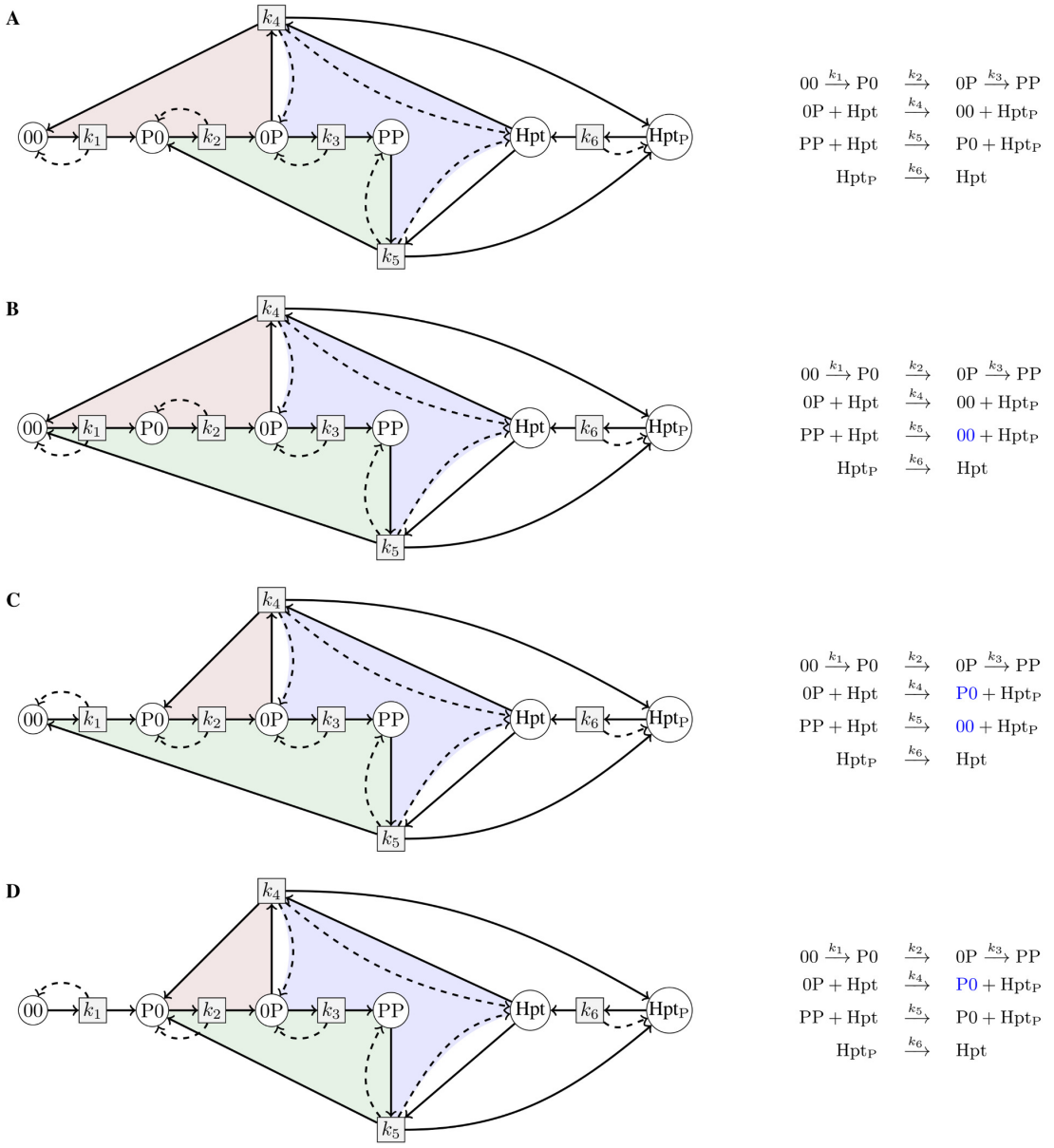


Figure 3.4: Bipartite directed species-reaction graphs for the minimal core system shown in Figure 3.2 (panel A) and several structural modifications of it (panels B-D). Modifications are highlighted in blue and do not necessarily result in biologically realistic systems. Despite containing the same number of feedback loops and components, the system shown in A is bistable, while those in B, C, and D are not. On each graph circular and rectangular nodes represent the biochemical species and reactions respectively. Dashed and solid edges indicate flows in the following manner: a solid edge is drawn from a species to a reaction to indicate that this species is consumed by that reaction, a solid (dashed) edge is drawn from a reaction to a species to indicate that this species is produced (consumed) as a result of this reaction. Three positive feedback loops are highlighted: the green and red positive feedback loops involve only solid edges, while the blue positive feedback loop involves combination of two dashed edges.

The aforementioned two feedback loops are complemented by a third feedback loop that becomes visible when we display the core model as a bipartite reaction graph

(figure 3.4). It is known that bistability requires at least one positive feedback loop in such a graph [129, 134] however, we find that alternative reaction schemes of the same size as the core system and implementing one or more feedback loops do not allow for bistability (figure 3.4). This shows that the reaction scheme in the core of the hybrid HK structure implements a particular, non-trivial mechanism for generating bistability. This mechanism is still intact in the full hybrid and unorthodox HK models, nested within a more complex reaction scheme that includes hydrolysis and reverse-phosphotransfer reactions. We find that these additional reactions allow tuning of the exact shape of the input-output response dynamics, with reverse phosphorylation providing the possibility of achieving more pronounced switch-like dynamics (figure 3.5). More broadly, we show that the mathematical findings for multistability extend to the full hybrid and unorthodox HK models, even when we take into account complex formation in the phosphotransfer reactions (see section 3.2.3).

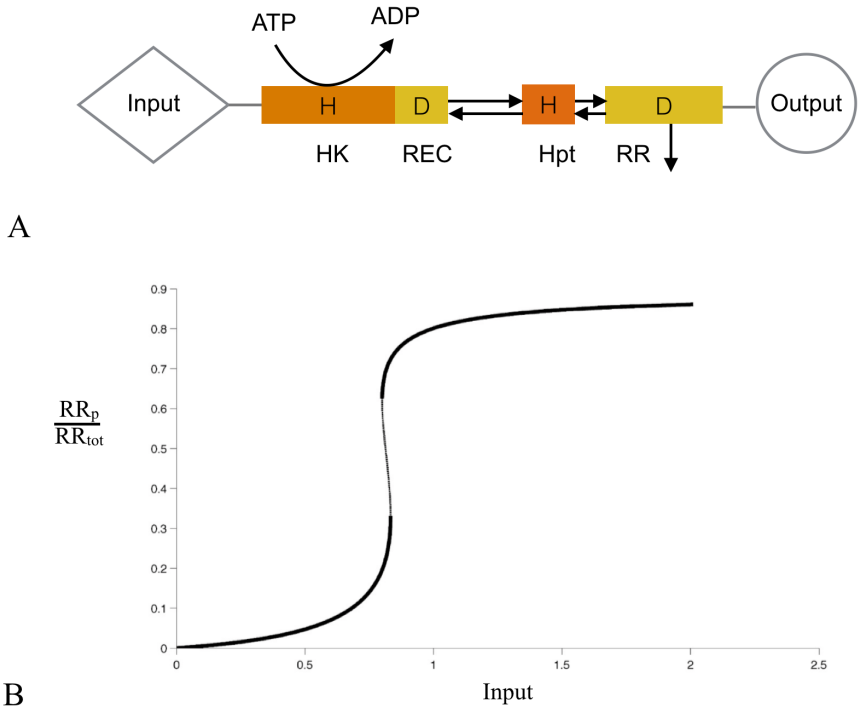


Figure 3.5: (A) Schematic representation of a full phosphorelay embedding a hybrid HK. The associated model for this system implements all reactions shown in equations (118)-(124). (B) Signal-response plot (i.e. dose-response) for the system in panel A, for a specific set of parameters (see Appendix F). The plot shows the fraction of phosphorylated RR at steady state while varying the auto-phosphorylation rate constants of HK, k_1 and k_3 (such that the ratio $\frac{k_3}{k_1}$ is fixed). The solid and dotted lines indicate stable and unstable steady states respectively.

3.3.2 Unbounded multistability and implementation of Boolean logic via component sharing

The key mechanisms for generating bistability in a single multidomain HK are the feedback loops among its internal phosphostates and the interlinkage of these to a downstream target. This rises the possibility that component sharing, in which several multidomain HKs share (i.e. phosphotransfer to) the same downstream target can lead to an increase in the number of steady states in the system. To address this possibility we analysed a generalised model of n HKs transferring phosphates to the same Hpt. We prove mathematically that such a system can attain $2n{+}1$ steady states under appropriate choices of parameters; to this end, we show that the steady states of a system comprising of n HKs transferring phosphates to the same Hpt are in correspondence with the positive roots of a polynomial of degree $2n{+}1$ in the concentration of phosphorylated Hpt (see sections 3.2.4 and figure 3.6). Of these steady states, n are proven to be unstable, and simulations show that the remaining $n{+}1$ steady states are, as expected, stable. Considering component sharing at the level of RR, we show that the system with m modules that phosphotransfer to the same RR, and where the i'th module comprises n_i hybrid HKs sharing a single Hpt, allows for $\prod_{i=1}^{m}(2n_i + 1)$ steady states (figure 3.4 D). In particular, the system comprising n phosphorelays, each consisting of a hybrid HK and a Hpt domain phosphotransferring to a common RR, can attain $3n$ steady states.

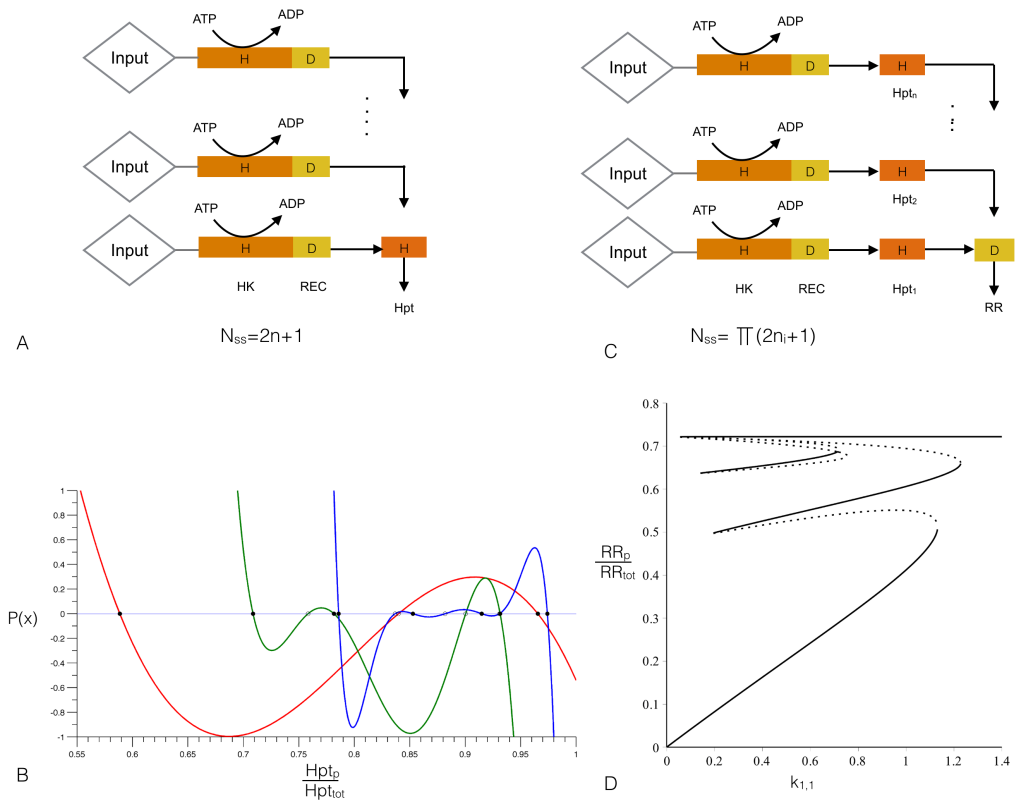


Figure 3.6: (A) & (C) Schematic representation of sharing of downstream components at the level of Hpt (panel (A)) or RR (panel (C)). (B) Plot of the polynomial function (see section 3.2.4) characterising the steady state level of the phosphorylated Hpt for the system shown in panel A for a specific set of parameter values (see section 3.2.5). Both the polynomial and the variable are scaled to the shown window. The polynomial is plotted in red, green, and blue for one, two, and three HKs respectively. Each crossing of the polynomial with the x-axis is a steady state of phosphorylated Hpt. Stable and unstable steady states are shown as filled and open circles respectively. (D) Signal-response plot for a system with 2 hybrid HKs, each with separate Hpts that share the same RR (as shown in panel (C), for $n = 2$). We assume that the auto-phosphorylation rate constants for HK1 and HK2 (when the REC site is unphosphorylated) are the same and determined by the same signal (i.e. $k_{1,1} = k_{2,1}$). The signal-response plot shows the fraction of phosphorylated RR at steady state for a given input level ($k_{1,1} = k_{2,1}$). The parameter values are chosen such that the system displays 9 distinct steady states (see section 3.2.5). The solid and dotted lines indicate stable and unstable steady states respectively.

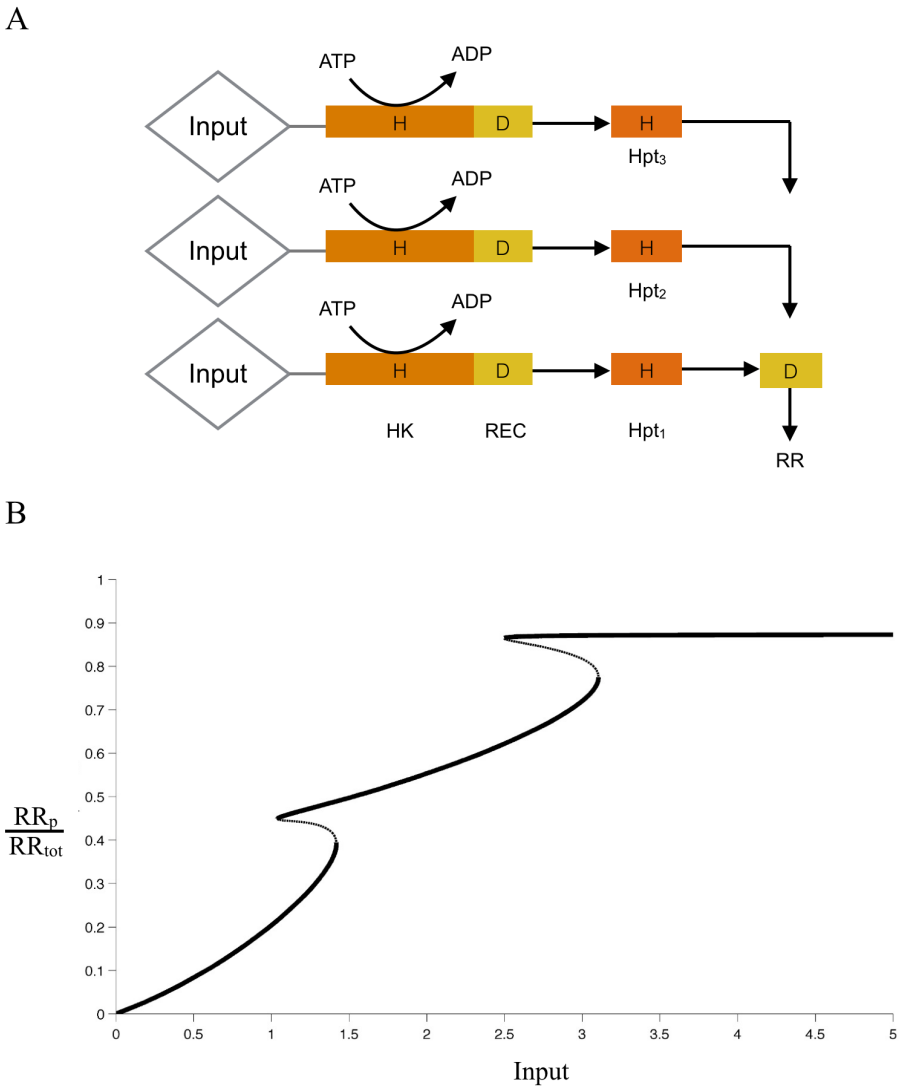


Figure 3.7: Schematic representation of a system with three hybrid HKs, each with their own Hpt that share the same RR (A). Sample signal-response plot for the system shown in panel (B). Assuming that the input level, that is, the phosphorylation rate constant from OO to PO, is the same for each hybrid HK (i.e. $k_{1,1} = k_{2,1} = k_{3,1}$) and the ratios $\frac{k_{1,3}}{k_{1,1}}, \frac{k_{2,3}}{k_{2,1}}, \frac{k_{3,3}}{k_{3,1}}$ are held fixed, the signal-response plot shows the fraction of phosphorylated RR at steady state for varying input level. The parameters are chosen (see Appendix F) such that the system displays bistability for different intervals of input levels. The solid and dotted lines indicate stable and unstable steady states respectively.

These mathematical proofs show that microbes can utilize individual hybrid and unorthodox HKs to implement multistability and are theoretically unbounded in their capacity to expand the number of available steady states through sharing of downstream components (Hpt or RR) among such HKs. We find that component sharing among multidomain HKs can also be utilized flexibly, and in other ways. For example, component sharing at the level of RR and using HKs sensing the same signal can be used to implement n bistable switches with distinct threshold signal

levels (figure 3.7). Perhaps more interestingly, HKs sensing different signals and component sharing at the level of RR can be used to implement Boolean logic gates. In particular, we could identify a simple architecture involving two HKs, sharing the same Hpt, that can implement an AND and an OR gate (see figure 3.8 & section 3.4.1). The system could be tuned between implementing these different logic gates simply by changing the total concentrations of components and the dephosphorylation rate of phosphorylated RR (see parameters listed in Appendix F). Furthermore, based on the above mechanistic understanding and parameter sampling, we could identify parameter combinations for the same system that allowed summation over the two signals (figure 3.8 B). This system response qualitatively matches the observations from the quorum sensing system of *Vibrio harveyi*, where 3 hybrid HKs share the same Hpt. Experiments with a modified version of that system, and involving just 2 HKs, have shown that the ability to perform a summation as shown in figure 3.8 is possible in a natural system.

3.4 Discussion & Conclusions

The ability to map environmental signals onto distinct internal physiological states or programmes is expected to be critical for single-celled microbes that often need to respond to signals arising from fluctuating environments and neighbouring populations. This physiological capacity usually requires signalling systems that can implement threshold dynamics or multistability. While previous studies have identified multi-site phosphorylation as a key biochemical mechanism to attain unbounded multistability, this mechanism is mostly lacking from microbial cells. Instead these cells rely on phosphotransfer reactions in two-component signalling for their environmental information processing. In this chapter, we mathematically prove that multidomain (hybrid & unorthodox) HKs found in these systems can display bistability by embedding multiple feedback loops within their own reaction scheme using fewer components in comparison to eukaryotic signalling networks like MAPK pathways [119].

When more than one multidomain HKs sharing the same downstream component, we find that a system can attain unbounded multistability. We derived several mathematical proofs relating the number of multidomain HKs sharing the same component, the position of the shared component in a phosphorelay and the number of steady states available to the system. These proofs extend to considering complex formation and show that microbes can attain unbounded multistability by employing two-component signalling. Furthermore, we find that the same principle of

component sharing among multidomain HKs can also be employed by the networks to implement Boolean logic gates when different HKs sense different signals.

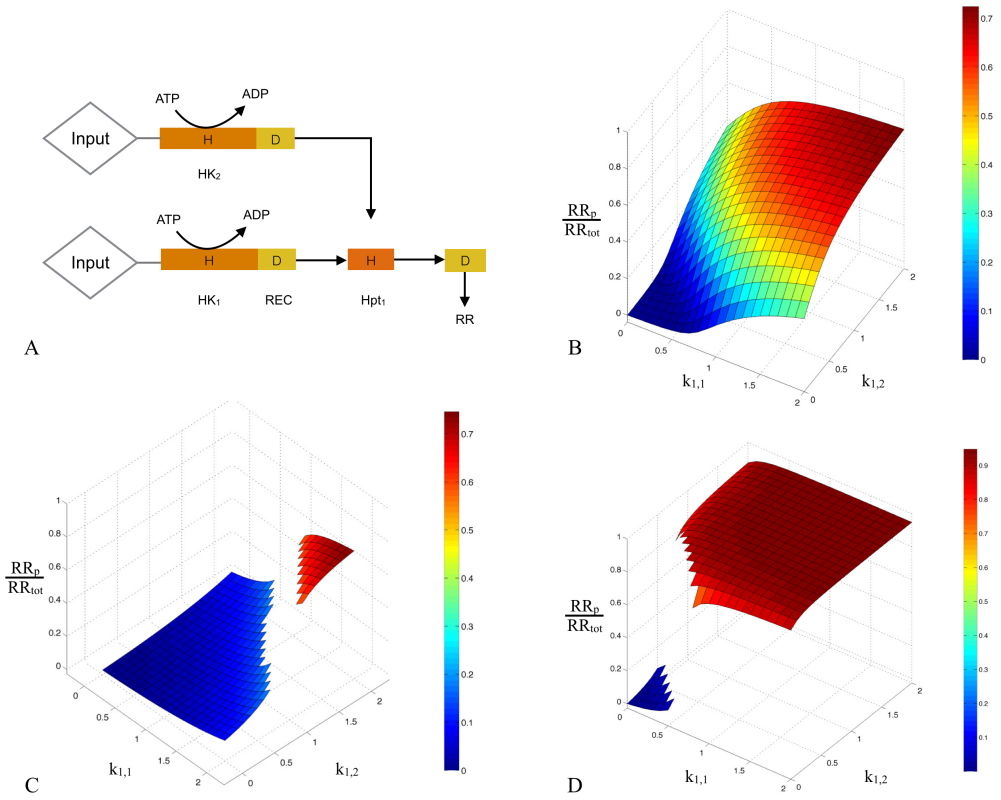


Figure 3.8: (A) Schematic representation of a system with two hybrid HKs, that share the same Hpt. (B - D) Implementation of different signal processing functions using the system shown in panel A; an "adder"as seen experimentally in *V. harveyi* [32] (**B**), a Boolean AND gate (C), and a Boolean OR gate (D). Each panel shows the fraction of phosphorylated RR at steady state for different auto-phosphorylation rate constants (i.e. input levels) acting on the two HKs. It is assumed that the two signals are specific for the two HKs ($k_{1,1}$,$k_{1,3}$ and $k_{2,1}$, $k_{2,3}$ respectively) and the ratios $\frac{k_{1,1}}{k_{1,3}}$ and $\frac{k_{2,1}}{k_{2,3}}$ are held fixed. The systems shown in panels C and D are bistable, with the blank regions of the surfaces corresponding to unstable regions. For the parameters used, see Appendix F.

3.4.1 Component sharing and Boolean Logic

In addition to unlimited multistability observed when networks share components, there could be a capacity for implementing logic functions in the resulting shared network. To determine whether reaction systems with multiple hybrid HKs can implement response dynamics mimicking Boolean logic, we examined how different HK-RR architectures respond to varying inputs to the system .[2] We sampled

[2]For each combination of signal level, the systems were run to steady state where a system was considered stable, only after simulating the ODEs for a sufficiently long duration and ensuring that the measured fluctuation in the system's output was below 1e-5 and the output RR_p levels have

different reaction rate constants and total concentration of proteins in the network to generate a steady state signal-response curve.

Representative parameter sets were found where these networks exhibited AND, OR and ADDER functions. The presented theoretical framework fits well with the few experimentally studied cases involving hybrid HKs. The quorum sensing system of *Vibrio harveyi* implements component sharing as discussed above. It has been experimentally shown that the system implements an 'ADDER 'function [134] that could be readily reproduced with the models presented here. Similarly, we found that the osmosensing system from yeast, implementing a hybrid HK [82], displays bistability and hysteresis under an experimentally measured and biologically feasible parameter regime. These analyses lend further support to the idea that the observed capacity for multistability arising from multidomain HKs is exploited by evolution and is implemented in natural two-component systems.

While systematic analyses in *Escherichia coli* and *Caulobacter crescentus* found mostly distinct HK-RR pairs [25, 135], a recent study suggest rapid diversification of RRs after duplication [6], these studies focused primarily on regular HKs. Where analysed, specific two-component systems involving multidomain HKs are found to display significant cross-talk [136] and also the exact type of component sharing described here (as seen in *Vibrio harveyi* [134]). The particular case of quorum sensing in *Vibrio harveyi*, for example, involves 3 hybrid HKs sharing the same Hpt and is indicated to perform a summation of at least two of the incoming signals [134] and results in similar dynamics to that shown in figure 3.8.

In this chapter, we therefore argue that component sharing among multidomain HKs could be seen as a design principle, which microbial cells can utilize flexibly to generate unbounded numbers of physiological steady states and implement logic in their operations.

converged to a steady state. All simulations were run using MATLAB and its native ode solvers (ode15s).

4

Biological significance of two component signalling and phosphorelay networks.

Chapter Overview

To understand the biological context of the dynamics embedded in different two component and phosphorelay architectures, we examine the distribution of different sensor proteins (single and multidomain sensor HKs) across prokaryotic genomes (and metagenomes) and annotate their partner response regulators (RR) to correlate the identified function, (regulated) metabolic process and location inside the cell of an RR with a network's dynamics defined by the sensor kinase present in the network. We examine the distribution of multidomain and regular HKs across genomes and study how they vary with respect to the size of a genome and a cell's external environment. We test if specific types of sensor kinase proteins are preferentially distributed in some organisms and examine if this depends on the size of a genome or the specific environment in which an organism occurs. For the former, we use genomic data available in P2CS, a two component protein database and for the later, we use a classification scheme that groups organisms as either extracellular, facultative intracellular or obligate

intracellular [137]. By identifying organisms from P2CS found in [137], we were able to examine the distribution of different types of HKs across species found in different environments. From these distributions, we were unable to find a clear evidence of correlation between genome size and the distribution of different HKs across genomes.

In addition to examining the distribution of different types of HKs, we examine the functional significance of sequences representing RRs partnered by different types of histidine kinases in a cell. We do this to find the physiological significance of different networks by identifying the functions that their response controls. We use BLAST2Go, a large scale functional annotation tool that identifies GO (Gene Ontology) IDs representing predicted metabolic processes, function and cellular locations of a response regulator protein using its sequence information and comparing with known sequences in NCBI's (nr i.e. non redundant) protein database. By annotating these response regulator proteins, we try to predict the function and metabolic processes that signalling networks containing these proteins regulate. Such an analysis allows us to predict the possible correlation between dynamics that different types of sensor kinases have a capacity for, along with the types of processes and functions that response regulator proteins regulated by them are involved in. This gives us an insight into the possible biological context in which theoretical findings regarding networks containing different proteins and architectures (as discussed in chapters 2 and 3.4) are relevant.

4.1 Introduction

Two component systems and phosphorelays are widely distributed across a large number of prokaryotes, archaea bacteria and in some eukaryotes. In genomes containing several of these networks, these signalling networks can be involved in regulating a number of different metabolic processes like sporulation (i.e. deciding when to form protective spores that ensure cell survival in the face of adverse environmental conditions) [3], osmoregulation (regulating internal osmotic pressure by controlling salt concentration) [108], cell cycle control (i.e. deciding when the cell divides and controlling the onset of different stages before the cell divides) [25], chemotaxis (regulating the movement of a microbial cell towards a food source by measuring the concentration of food in the environment) [10], and quorum sensing [4]. In each of these roles, networks have different architectures and corresponding response dynamics.

For example, processes like sporulation in *Bacillus subtilis* [3] embed a basic

tunable architecture (discussed in [85]) when regulating a cell's decision to form spores under environmental stress. This network integrates signals from quorum molecules (signalling molecules that pass information between cells in a microbial community) in the environment with signals of stress to calculate a food per cell estimate that decides when the organism forms spores for survival. The nature of this architecture is such that the response from this network can be tuned via transcriptional feedback and dephosphorylation of Spo0F-p by Rap-A, a protein responsible for integrating information from quorum signals into this pathway. In ([17, 105, 106]) we see signalling networks where ligand binding at a bifunctional histidine kinase toggles its kinase/phosphatase activity and affects the overall response characteristics of the network depending on the surrounding external environment. In *Vibrio harveyi*, we find an ADDER function implemented by LuxN and LuxPQ to control gene expression through a synergistic detection of multiple inputs. In Ann *et.al.* [82], we see that Ypd1 dependent phosphotransfer reactions display bistability and hysteresis under an experimentally measured and biologically feasible parameter regime when regulating the osmotic pressure in the cell. All these examples lend support to the idea that different dynamics emerging from different signalling architectures are exploited in cells to uniquely suit regulatory requirements of specific metabolic processes that the network regulates. There could also be additional environmental dependence on these networks carrying out specific functions in the cell using a specific signalling architecture.

In previous chapters, we discussed phosphorelay architectures and their underlying reaction schemes which embed tunability and multistability. We examined the significance of different phosphorelay architectures and found that an interplay between forward and reverse phosphotransfer along with a presence of hydrolysis reactions at specific proteins in the network is essential for tunability in a network, where the response from the same network can be either sigmoidal or hyperbolic, depending on reaction parameters and total protein concentrations. Building on this, by examining different types of histidine kinases (multidomain and single domain) and their effects on the overall dynamics of a signalling networks we showed that phosphorelays with multidomain proteins have a capacity for multiple steady states due to the double negative feedback loops occurring in their chemical reaction networks. We showed that a single additional phosphorelay domain on the sensor HK (hybrid HK) and a downstream phosphate receiver domain can embed reactions required in a system for it to be multistable. We also found that networks with shared signalling architectures involving multidomain proteins have a capacity to exhibit unlimited multistability and execute logical operations like AND, OR and ADDER ([138]).

Models built in chapters 2 and 3 mainly examine the theoretical signal processing capabilities of different signalling networks. To understand more about the biological significance of these findings, we examine the genomic distribution of different types of histidine kinases and predict the function, cellular location and metabolic processes involving response regulators partnering these HKs. By examining the distribution of multidomain HKs with respect to regular HKs across genomes, we explore if this distribution depends on the size of the genome. In addition to this, with multistability in a network enabling a cell to achieve multiple steady states, we test if prokaryotes with smaller genomes and especially those which occur in highly variable environments have more multidomain proteins than regular HKs to leverage multistability to perform more complex information processing using fewer resources [139].

Using P2CS, a two component signalling database [140, 141] and environmental classification defined by Bordenstein *et.al.* [142], we examine the distribution of different types of HKs across genomes and also study the environmental impact on the observed distribution of different types of HKs.

4.2 Understanding the biological context of signalling networks

Let us consider the examples of *Treponema primitia ZAS-2* and *Streptococcus pyogenes*. *Treponema primitia ZAS-2*, has three times more multidomain HKs [126] than regular HKs and *Streptococcus pyogenes*, does not have any multidomain HKs and contains only regular histidine kinases. When we compare the environment in which the two species occur, we see that *Treponema primitia ZAS-2* is a spirochete from the gut microbiota while *Streptococcus pyogenes* is a human pathogen which infects the respiratory tract and superficial layers of the skin. The environments in which these organisms occur and the distribution of different HK types in these genomes, coding density[1] and cross talk (triad, tetrad, pentad genes) between the TCS genes (see table 4.1) are all part of what we define here as the biological context for a signalling network in a cell. The nomenclature of triad, tetrad and pentad genes refers to the number of two component genes found in the same gene cluster on the genome. This hints at the possible presence of branched signalling pathways

[1]The term coding density represents the percentage of genes in a genome which produce a functional product.

with sharing of components during signal transfer between different two component proteins.

Species Name	*Streptococcus pyogenes HSC5*	*Treponema primitia ZAS-2*
Total Genome Size	1818351	4059867
Coding Density	84.77	89.81
Average Gene Length	884	1035
Regular HK	12	11
Hybrid HK	0	16
Unorthodox HK	0	17
Orphan Genes	3	64
Paired Genes	11	16
Triad Genes	0	1
Tetrad Genes	0	1
Pentad Genes	0	1
Hexad Genes	0	0
Heptad Genes	0	0
Multidomain HK/Regular HK	0	3
Environment	respiratory tract / skin	termite hindgut

Table 4.1: Genomic information for *Streptococcus pyogenes HSC5* and *Treponema primitia ZAS-2* from P2CS.

Table 4.2 shows a list of the top 20 genomes with more multidomain HKs than regular HKs and table 4.3 shows the top 20 genomes with more regular HKs than multidomain HKs. To understand if there is a broad correlation between genome size and a selective distribution of different types of sensor kinases, we extract genomic data for TCS proteins across all prokaryotic genomes listed in P2CS.

Species Name	Total Genome Size (bp)	Coding Density	Average Gene Length	Regular HK	Hybrid HK	Unorthodox HK	Multidomain HK/Regular HK	Environment
Treponema primitia ZAS-2	4059867	89.81	1035	11	16	17	3	Termite Gut
Fibrobacter succinogenes subsp. succinogenes S85	3843004	84.03	1125	3	6	1	2.3333	Rumen
Treponema azotonutricium ZAS-9	3855671	90.73	1007	16	26	11	2.3125	Termite Gut
Magnetococcus sp. MC-1	4719581	86.5	1099	30	47	17	2.1333	Estuary in Minessota (magnetotactic bacteria)
Sphingobium chlorophenolicum L-1	1368670	87.28	1082	3	6	0	2	Soil (Can mineralize PCP pesticide)
Bacteroides vulgatus ATCC 8482	5163189	88.48	1124	17	29	0	1.7059	Small intestine of humans
Cyanobacterium aponinum PCC 10605	4114099	80.19	982	16	22	2	1.5	Fresh Water
Allochromatium vinosum DSM 180	3526903	90.66	1047	27	16	20	1.3333	Sandy beaches (soil) purple sulphur bacteria
Synechococcus sp. PCC 7502	3510253	83.94	907	18	22	2	1.3333	Marine environment
Coxiella burnetii Dugway 5J108-111	2158758	82.04	889	3	3	1	1.3333	Obligate intracellular, eukaryotic phagocyte
Prevotella dentalis DSM 3688	1890695	87.86	1177	3	4	0	1.3333	Gut
Cyanothece sp. PCC 7425	5374574	85.78	928	38	46	2	1.2632	Rice fields (nitrogen fixing)
Bacteroides xylanisolvens XB1A	5976145	80.72	1095	24	30	0	1.25	Human gut
Capnocytophaga ochracea DSM 7271	2612925	86.28	1038	4	5	0	1.25	Oral cavity
Chlorobium phaeovibrioides DSM 265 (Prosthecochloris vibrioformis)	1966858	91.2	1023	4	4	1	1.25	Acquatic environment (green sulphur)
Ramlibacter tataouinensis TTB310	4070193	92.41	969	36	39	5	1.2222	Sand
Isosphaera pallida ATCC 43644	5472964	84.38	1252	14	15	2	1.2143	Hot spring (phototactic)
Aliivibrio salmonicida LFI1238	1206461	77.18	950	5	4	2	1.2	Marine pathogen (salmon)
Opitutus terrae PB90-1	5957605	88.56	1144	54	62	2	1.1852	Rice paddy soil
Alistipes finegoldii DSM 17242	3734239	86.45	1038	9	10	0	1.1111	Human appendix

Table 4.2: Top 20 P2CS genomes with more multidomain than regular HKs.

Species Name	Total Genome Size (bp)	Coding Density	Average Gene Length	Regular HK	Hybrid HK	Unorthodox HK	Multidomain HK / Regular HK	Environment
Streptomyces griseus subsp. griseus NBRC 13350	8545929	88.12	1055	82	1	0	0.01	soil-inhabiting, filamentous bacteria (produces antibiotics)
Stackebrandtia nassauensis DSM 44728	6841557	90.95	975	77	1	0	0.01	soil
Conexibacter woesei DSM 14684	6359369	93.96	1010	59	1	0	0.02	soil
Kribbella flavida DSM 17836	7579488	89.7	979	59	1	0	0.02	soil, potato, alum slate mine, patinas of catacombs or from horse racecourses
Bacillus thuringiensis serovar finitimus YBT-020	5355490	84.29	829	52	1	0	0.02	insect
Nocardiopsis alba ATCC BAA-2165	5848211	86.26	916	51	1	0	0.02	honeybee gut
Bacillus anthracis str. Sterne	5228663	84.11	831	50	1	0	0.02	N.A.
Streptomyces coelicolor A3(2)	8667507	88.76	994	98	2	0	0.02	soil (can produce antibiotics, immunosuppressants)
Microlunatus phosphovorus NM-1	5683123	90.93	968	47	0	1	0.02	EBPR activated sludge
Streptomyces davawensis JCM 4913	9466619	90.12	1003	92	2	0	0.02	soil
Amycolatopsis orientalis HCCB10007	8948591	90.32	996	91	2	0	0.02	(produces vancomycin)
Desulfitobacterium dehalogenans ATCC 51507	4321753	84.4	909	45	0	1	0.02	freshwater pond (Athens, GA)
Streptomyces hygroscopicus subsp. jinggangensis 5008	10145833	83.13	953	86	2	0	0.02	Jinggang Mountain area of China (antibiotic producer)
Streptomyces hygroscopicus subsp. jinggangensis TL01	9840102	83.96	959	85	2	0	0.02	N.A.
Clostridium kluyveri NBRC 12016	3896121	83.39	937	40	1	0	0.03	First isolate mud of a canal in Delft
Streptomyces sp. PAMC26508	7526197	88.43	955	75	2	0	0.03	soil
Mycobacterium vanbaalenii PYR-1	6491865	91.52	994	37	1	0	0.03	first isolated from petroleum-contaminated estuarine sediments
Streptomyces fulvissimus DSM 40593	7905758	88.84	1014	74	2	0	0.03	N.A.
Rhodococcus jostii RHA1	7804765	91.23	987	36	1	0	0.03	soil contaminated with lindane
Rhodococcus pyridinivorans SB3094	5227080	88.61	966	36	1	0	0.03	N.A.

Table 4.3: Top 20 P2CS genomes with more regular than multidomain HKs.

4.2.1 Distribution of regular and multidomain HKs across prokaryotic genomes in P2CS

P2CS, is a comprehensive repository of two component signalling systems, and contains data from 1125 (July, 2013) prokaryotic and archaea bacterial genomes, along with metagenomic data. We consider only 766 out of 1125 genomes as there were several species with multiple subspecies included in the database. For these organisms, we considered a single representative entry and exclude the others as they had similar genomic composition. For these 766 species, we extracted information on the number of different TCS genes, predicted signalling partners (includes HK-RR pairs and multiples where more than one ($n = 1, 2 \ldots 5$) HKs partner with an RR or *vice-versa*) and database identifiers (GI IDs) for nucleotide and amino acid sequence records for TCS genes in NCBI's [143] protein database.

We extract the number of different types of HKs found across all prokaryotic genomes in P2CS, using raw data provided by Phillip and Barakat [140, 141] (via personal correspondence) and by writing Perl scripts to directly fetch missing data from `http://p2cs.org`. Using this data, we plotted the number of regular vs multidomain HKs and the ratio of multidomain HKs vs total genome size (panels A and B respectively in figure 4.1). The points highlighted in red (figure 4.1, panel B) represents genomes where the ratio of multidomain to regular HK is > 1. From (figure 4.1A), we do not see a meaningful correlation between the occurrence of regular and multidomain HKs across genomes and in figure 4.1B, we find very few genomes (40) with more multidomain HKs than regular HKs. When we examine organisms with small genomes (< 1 Mbp[2]), i.e. those that fall in regions R1, R2 in the plot, there are very few genomes with a ratio of multidomain HK to regular HK > 1. Figure 4.2 shows that nearly 70 % of the genomes smaller than 1 Mbp have no multidomain HKs, 13% have more regular HKs than multidomain HKs and about 17% have more multidomain HKs than regular HKs. If we consider genomes with sizes between 1 and 3.8 Mbp (the average genome size of of all genomes in P2CS), we find that 204 out of 399 genomes included in the dataset, i.e. 51% of the genomes, have no multidomain HKs at all, 46% have more regular HK than multidomain HK and the remaining 3% of the genomes have more multidomain HKs. 70% of the genomes (< 1 Mbp) and 51% of the genome (> 1 Mbp & < 3.8 Mbp) have no multidomain proteins, and the remaining genomes (> 3.8 Mbp) shows a clear majority presence of regular HKs (nearly 97%).

[2]Mbp stands for Mega base pair i.e. 10^6. It indicates the length of the DNA strand and represents the number of base pairs in the genome.

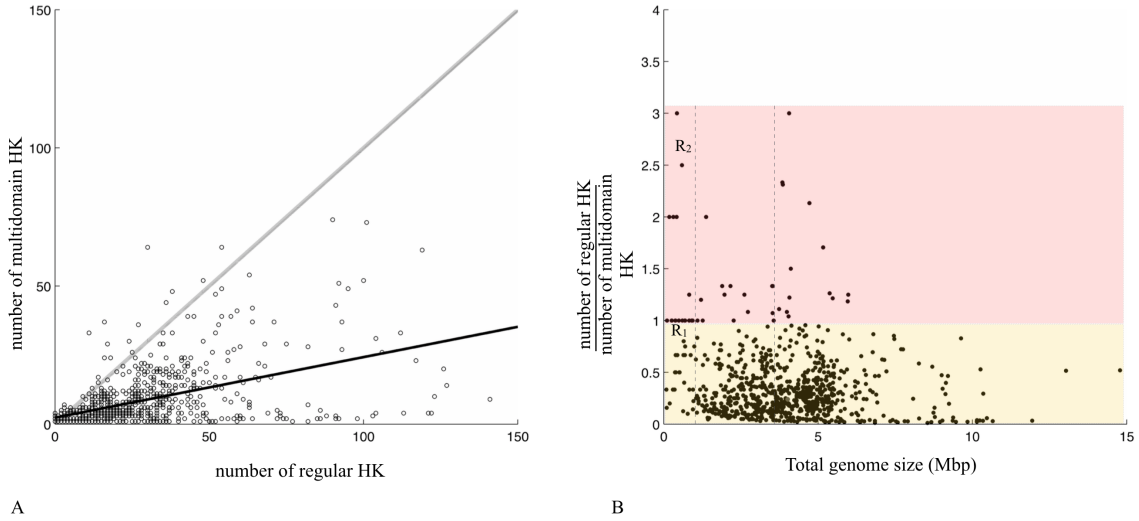


Figure 4.1: Shows the number of regular and multidomain HKs on the x-axis and y-axis respectively. The dotted line is the reference correlation line for perfect positive correlation ($R = 1$) and the continuous line represents the actual positive correlation of ($R = 0.49$) between the number of regular and multidomain HKs (A). Shows the ratio of multidomain to regular HKs plotted against total genome size of the organisms corresponding to that ratio. Data points with values > 1 on the y-axis are highlighted in red and represent genomes that harbour more multidomain than regular HKs. Genomes where there were no multidomain proteins and/or no regular HKs are not included in this plot (B).

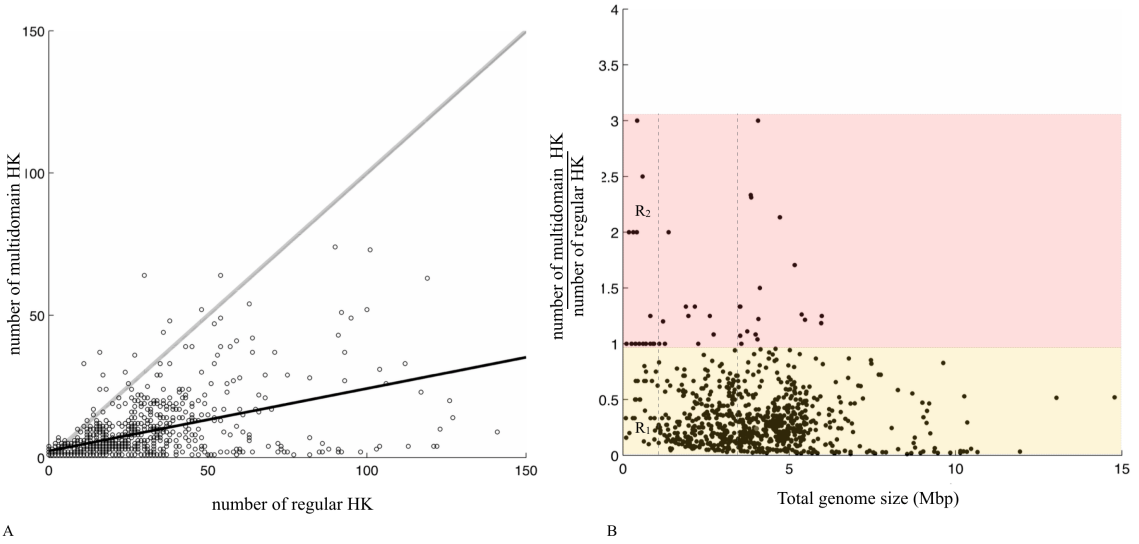


Figure 4.2: Plot (A) shows the distribution of regular and multidomain HKs amongst genomes in P2CS with a genome sizes < 1 Mbp. The plot shows data from each genome represented as a data point, with the x-axis and y-axis displaying regular and multidomain HKs respectively. Data points with values > 1 on the y-axis are highlighted in red and represent genomes that harbour more multidomain than regular HKs. (B) Histogram representing the distribution of the ratio of multidomain HK vs regular HK (x-axis) and the number of genomes (y-axis).

As prokaryotic species occur in a variety of environments at different stages in their life cycle, we carry out a similar examination of a subset of organisms considered in figure 4.2 and study the genomic distribution of different sensor HKs in specific environments to test if organisms occurring in highly variable environments have more multidomain than regular HKs.

4.2.1.1 Distribution of regular and multidomain HKs in specific environments

Previous studies [126, 144–149] on the evolution of two component systems have shown that the evolution of different prokaryotic signalling architectures depends on gene duplication, separation of domains and introduction of additional intermediate domains on the same protein [126] via horizontal gene transfer and lineage specific expansion. With organisms living in different environments at different stages in their life cycle, different mechanisms are employed to evolve network architectures that enable them to cope with their surrounding environment and regulate their internal physiological state appropriately. In this section, we will examine a subset of genomes (see figure 4.1) occurring in different environments (as characterised in [142]) and study their sensor HK distribution patterns. Depending on the environment in which different stages of their life cycle occurs, organisms can be classified as extracellular, facultative intracellular or obligate intracellular. Extracellular organisms are those which reside outside a living host cell in all stages of their life cycle. Facultative intracellular organisms are those which have parts of their life cycle inside a host and some parts where they are outside the host cell. Obligate intracellular organisms are those which only survive inside a host cell. Each of these environments have different characteristics. Organisms occurring outside the cell, i.e. extracellular, face a highly variable environment while facultative intracellular organisms have stages of their life cycle where they are outside the cell, facing a highly varying environment and other stages inside the cell where they are in a more stable homoeostatic environment. For organisms which are obligate intracellular, they are always found inside a host cell, in a stable, less variable environment. Unlike facultative intracellular organisms, obligate intracellular organisms are always inside a host and embed mechanisms that integrate themselves within the host cell machinery. Depending on the environment, organisms evolve different strategies for survival and can employ properties like a capacity for crosstalk and multistability (chapter 3) [138] for processing multiple inputs and employ computational properties observed in some signalling architectures (as discussed in chapter 3) implementing logic functions like AND, OR and ADDER.

Of all the 384 organisms which were included in the study by Bordenstein *et.al.*,

we found 124 extracellular, 24 facultative intracellular and 5 obligate intracellular organism entries in P2CS. From these entries, we extract information regarding the number of different types of HKs found in these organisms and generate the plots shown in figure 4.3. From figure 4.3B, we see that there are very few genomes with more multidomain HKs than regular HKs and in figures 4.3C and 4.3E, it can be seen that, there are no organisms with more multidomain HKs than regular HKs. This plot also shows limited representation of organisms characterised as facultative intracellular and obligate intracellular and there isn't sufficient information for us to examine if organisms occurring in different environments have a specific distribution of different sensor histidine kinase proteins.

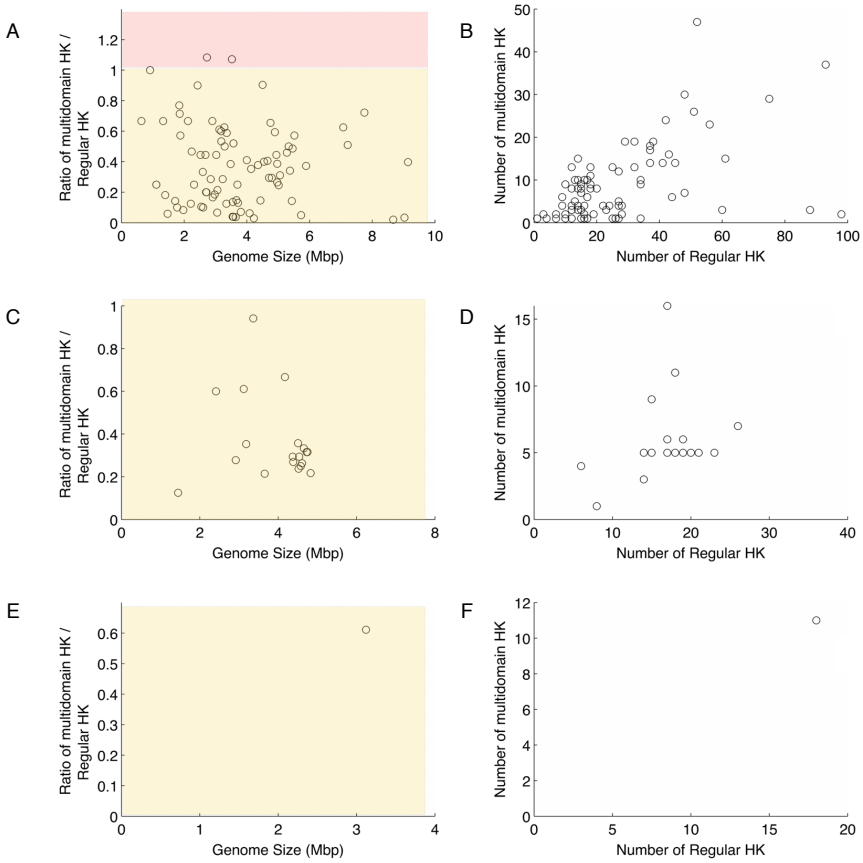


Figure 4.3: The genomic distribution of different HK types across three groups of organisms classified on the basis of their external environment. Panels (A-B) represent extracellular organisms, (C-D) represents facultative intracellular organisms and (E-F) represent obligate intracellular organisms. Panels (A,C,E) show the ratio of multidomain HKs vs regular HKs (y-axis) to total Genome Size (x- axis) and (B,D,F) show the number of multidomain HKs (y-axis) vs the number of regular HKs (x- axis). In panel A, red colour has been used to highlight genomes where the ratio of multidomain HKs vs regular HKs is greater than 1. Genomes where there were no multidomain proteins and no regular HKs have been removed from the plot.

Even as [142] considered 384 organisms, only 153 out of those organisms had

corresponding data in P2CS (as of July 2013). In this study, we are using data only from P2CS as it has a high accuracy of predicting proteins [141] found in two component and phosphorelay networks.

In addition to examining the distribution of different types of histidine kinase proteins across genomes and environments, we examine the signal transducer i.e. response regulator proteins that these histidine kinases partner with. Analysing them can help in identifying the metabolic processes, and functions that different HKs regulate through their partner sensor HKs.

4.2.2 Functional Annotation of RR

To carry out a functional annotation of the RR proteins in the network, we use a high throughput sequence annotation tool called BLAST2GO. Annotation of a (gene or a protein) sequence is a process in which we identify the location (within the cell), function (of the gene/protein) in the cell and the metabolic process involving the protein (or gene). BLAST2GO [150–152], is a high throughput functional annotation platform which can be used to identify the function, location and process involvement of protein sequences representing all RRs identified as partners for different HKs found across genomes in P2CS ([153]). It uses a homology based approach where it identifies target sequences (with known function, metabolic process involvement and location) similar to the query sequence and derives annotations for the query sequence based on its similarity to a target sequence. By identifying multiple sequences which are similar to the query sequence and using an annotation scoring method, BLAST2GO derives the most appropriate ontology (GO: Gene Ontology) IDs that represent the process, function and cellular location of the query sequence.

Gene ontology IDs represent the detectable attributes of a gene (or protein) and are derived from a consensus database of terms representing different attributes of a gene (or a protein) derived from sequences found across all species. This database maintains a vocabulary of function, process and location of all genes and gene products (proteins) in the cell. Each GO ID has a term name, an alphanumeric identifier of the form GO:$\mathrm{X_1X_2X_3X_4X_5X_6X_7}$, where every X_{i} ($i = 1 \cdots 7$) can be any number between 0 and 9, a citation for the source, and a namespace definition which represents the domain of the term (i.e. molecular function, process and cellular component). These terms annotate sequences and identify the role they play in a cell.

In this study, we download protein sequences for RRs regulated by regular, hybrid

and unorthodox HKs. Regular HK mediated RR sequences are categorised as group-I and RRs mediated by hybrid and unorthodox HKs are classified as group-II (see table 4.4). Sequences in these two datasets were separately annotated and different stages in their annotation process are explained below. We begin by first extracting protein sequences which will be annotated.

4.2.2.1 Extracting sequences from P2CS

A list of GI IDs [3] for all HKs found in P2CS, and their corresponding RR partners were extracted using the raw data provided by Phillip and Barakat (personal correspondence) and by writing Perl scripts that scanned (web-scraping scripts) `http://p2cs.org` and extracted GI IDs representing HKs and RRs on P2CS. EGquery and EFetch utilities within the E-utilities programme suite at NCBI (`http://www.ncbi.nlm.nih.gov/books/NBK25497/`) were also used along with writing another set of Perl scripts to query the protein database on NCBI and download all the necessary protein sequences as a single batch.

	# sequences
Group I (Regular HK mediated RR)	27095
Group II (Hybrid and multidomain HK mediated RR)	2357

Table 4.4: The number of RRs downloaded from NCBI (July, 2013). Group I: RRs regulated by regular HKs; and Group II: RRs regulated by hybrid and unorthodox HKs.

The first step in the annotation process using BLAST2GO involves running a BLAST [154] search to find other proteins which are similar to the query sequence.

4.2.2.2 BLAST

BLAST stands for Basic Local Alignment Search Tool and as part of this, each query sequence is compared with all the sequences in the protein nr (non redundant) database one at a time and the top 20 similar sequences are identified [154]. The results from this blast search for all the sequences are compiled and stored as an Xml file. The top BLAST hits ([155]) are then passed on through to the next step, where appropriate GO IDs are assigned to sequences.

[3] GI ID stands for Genome Identifier IDs and represent the identifier used to refer to sequence records in the NCBI database.

4.2.2.3 GO mapping

Once the top 20 target sequences similar to all query sequences are identified, BLAST2GO uses information available on public resources like NCBI [143], PIR (`http://pir.georgetown.edu/`) and GO (`http://geneontology.org/`) to link short-listed BLAST sequences (for every query protein) to GO IDs found in the Gene Ontology database [155, 156]. This database contains millions of GO annotated sequences. Each sequence can be assigned three different types of GO IDs corresponding to the three attributes: process, component (location) and function.

4.2.2.4 Annotation

BLAST hits mapped to their GO-IDs ([155]) along with the query sequences are passed on through to the annotation step, where using GO ID mappings on the target hit sequences, BLAST2GO predicts GO IDs for every query sequence. The BLAST2GO annotation algorithm examines the similarity between query and hit sequences, the quality of the source of initial GO assignments to the target hit sequences, and the structure of the GO Directed Acyclic Graph (DAG) [150] representation of the context of biological dependences mapped using these GO IDs. From this, an annotation score (AS) is computed ([150]) using the formula

$$DT = max(similarity \cdot EC_{weight})$$

$$AT = (\#GO - 1) \cdot GO_{weight}$$

$$AS = DT + AT \tag{185}$$

The AS is composed of a direct term (DT) that represents the highest similarity value among the hit sequences bearing this GO term, weighted by a factor corresponding to its evidence code (EC).[4] EC_{weight} used to calculate DT is an indication of the reliability of the GO annotation. A GO term EC is present for every annotation in the GO database and indicates the procedure of functional assignment. ECs vary depending on the inferred mechanism, for example-Inferred from Direct Assay (IDA) to unsupervised assignments such as Inferred by Electronic Annotation (IEA). The second term (AT) introduces the possibility of abstraction into the annotation algorithm. Abstraction is defined as the annotation to a parent node when several child nodes are present in the GO candidate pool.

[4]EC or Evidence Code: All annotations are associated with an Evidence Code which provides information about the quality of the functional assignment of the annotation, depending on the method used to derive the annotation `https://www.blast2go.com/images/b2g_pdfs/b2g_user_manual.pdf`.

AT multiplies the number of total GOs unified at the node by a user defined GO_{weight} which controls the possibility and strength of abstraction. The expression for AR shows how we select the lowest term per branch that lies over a user defined threshold [151]. The GO distribution seen in figure 4.4 gives us an indication of the number of GO IDs assigned to sequences in both group-I and group-II.

In this step, we set the E-value filter as 1e-6 and an annotation cut-off of 60. A GO weight of 15 is used along with a Hsp-Hit coverage of 80. The default evidence code weights set-up in BLAST2GO were retained and sequences from groups I and II were annotated.

4.2.2.5 Comparing multidomain HKs with regular HKs

Once we derived the annotations for both regular and multidomain HKs (see table 4.4), we examined the annotations assigned to both groups and used enrichment analysis[5] [152] (`http://resources.qiagenbioinformatics.com/manuals/isv/Blast2GO/Blast2GO_PRO_User_Manual.pdf`) along with Fischer's exact test using a robust False Discovery Rate (FDR) correction to asses the functional differences between sets of functional annotations assigned to groups I and II. This was carried out by BLAST2GO using the Gossip package ([157]) for statistical assessment of differences in GO term abundance between two sets of sequences. We loaded all the sequences found in Groups I and II (table 4.4) along with their annotations and assigned sequences from Group II (multidomain) as a test set and sequences in group I as the reference dataset. We then carried out a single-tailed analysis and focussed on the well represented functions in these test groups, ignoring the under represented annotations. By carrying out a positive enrichment analysis (with a false discovery rate (FDR) threshold of 0.05), we derived a list containing significant GO terms ranked according to their representation in the datasets (figures 4.5, 4.6 & 4.7). This gives us an insight into physiological differences between single domain regular HKs and multidomain HKs.

[5]Enrichment analysis is an approach used to compare two different sets of genes or proteins with annotations to identify significant over or under represented functions assigned to them.

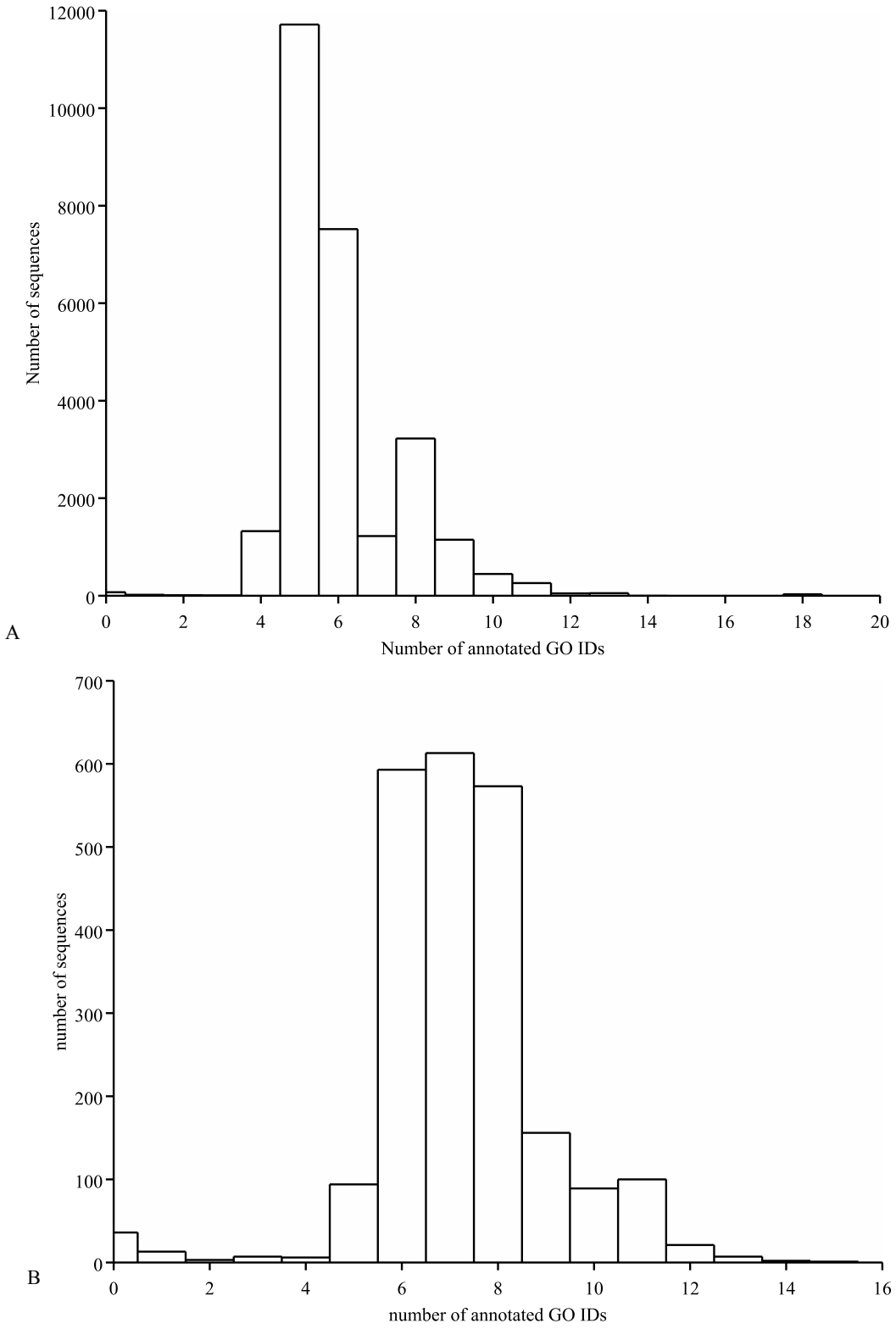


Figure 4.4: Shows the number of GO IDs annotated for query sequences from group I (A) & group II (B) (table 4.4).

GO Annotation: Cell component (Location)

From figure 4.5, we see that a significant number of Group II proteins have predicted location inside the cell with the intracellular annotation (GO: 0005622 http://amigo.geneontology.org/amigo/term/GO:0005622) indicating that RRs in group II are found in locations within (but not including) the plasma membrane, and excluded large vacuoles and masses of secretory or ingested material. In addition to this more than 32% of sequences in group II have a predicted association with a macromolecular complex (GO:0032991, http://amigo.geneontology.org/amigo/term/GO:0032991) and protein complex (GO:0043234, http://amigo.geneontology.org/amigo/term/GO:0043234) while $< 2\%$ of group I sequences have the same association. Around 17 % of the annotated sequences in group II are also associated with transcription factor complex (GO:0005667, http://amigo.geneontology.org/amigo/term/GO:0005667).

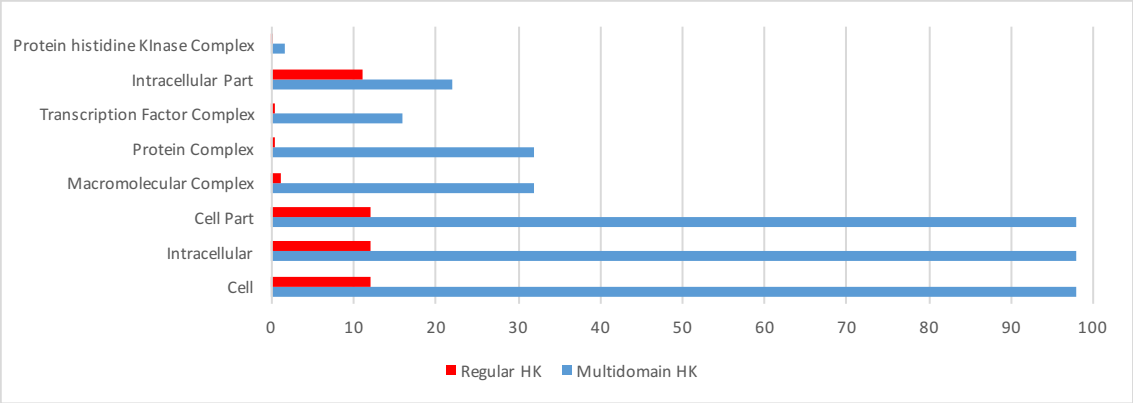


Figure 4.5: Shows results from the analysis carried out to ascertain the statistically significant functional differences between groups I and II. The y-axis shows significantly enriched GO terms that represent the cellular locations of annotated proteins and the x-axis gives the relative frequency of each term (i.e. % of all the sequences in the group which were mapped to the GO ID represented by this GO term). Red bars correspond to group I (regular) and blue bars correspond to group II (multidomain HKs).

GO Annotation: Molecular Function

When we examine the predicted molecular functions for annotated sequences in group I and II, we find that a significant proportion of the sequences in both groups are predicted to be involved in ion binding (GO:0043167, http://amigo.geneontology.org/amigo/term/GO:0043167) and catalytic activity(GO:0003824, http://amigo.geneontology.org/amigo/term/GO:0003824). The other molecular functions involving both groups are transferase ac-

tivity (GO:0016740, http://amigo.geneontology.org/amigo/term/GO:0016740), phosphorous-oxygen lyase activity (GO:0016849, http://amigo.geneontology.org/amigo/term/GO:0016849) and lyase activity (GO:0016829, http://amigo.geneontology.org/amigo/term/GO:0016829) (See figure 4.6).

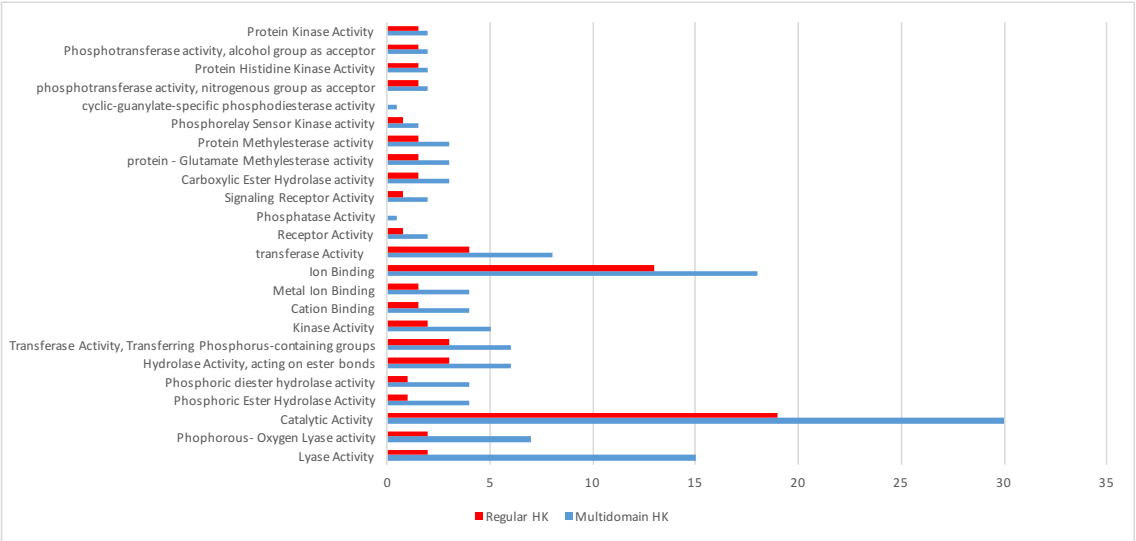


Figure 4.6: Shows results from the analysis carried out to ascertain the statistically significant functional differences between groups I and II.The y-axis shows significantly enriched GO terms that represent the molecular function that the annotated proteins are involved in and the x-axis gives the relative frequency of the term (i.e.% of all the sequences which were mapped to this GO ID represented by this GO term). Red bars correspond to group I (regular HKs) and blue bars correspond to group II (multidomain HKs).

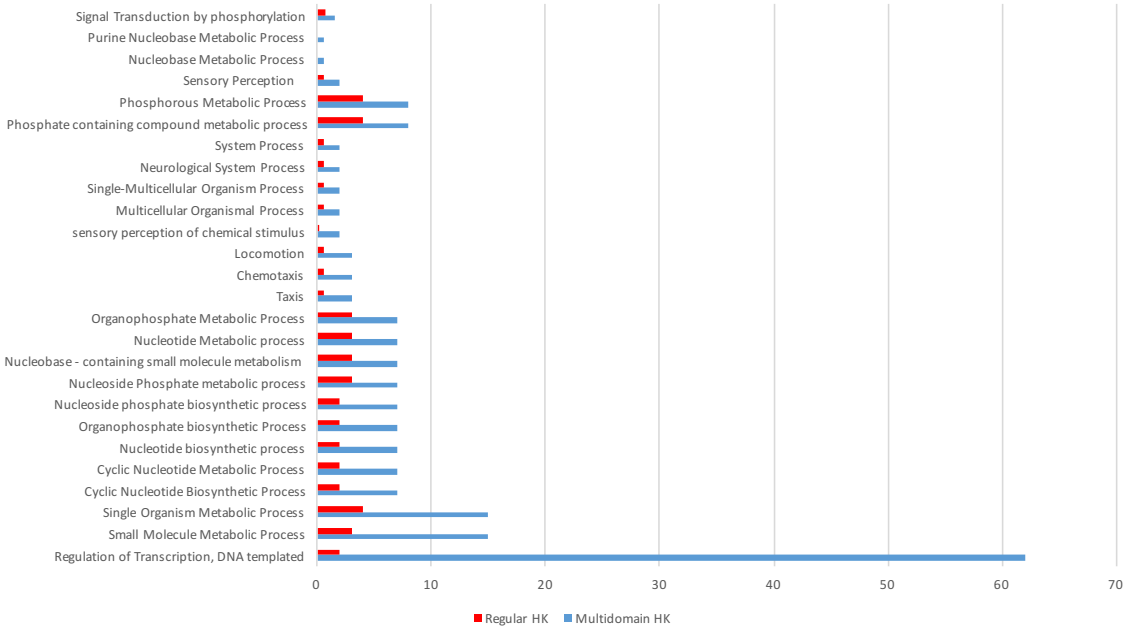


Figure 4.7: Shows results from the analysis carried out to ascertain the statistically significant functional differences between groups I and II. The y-axis shows significantly enriched GO terms that represent the cellular processes that the annotated proteins are involved in and the x-axis gives the relative frequency of the term (i.e. % of all the sequences which were mapped to this GO ID). Red bars correspond to group I (regular HKs) and blue bars correspond to group II (multidomain HKs).

GO Annotation: Metabolic Processes

Based on the predictions made regarding the metabolic processes that these sequences are involved in (see figure 4.7), multidomain HK regulated RR proteins in group II appear to have more than 60% of the sequences annotated with a predicted role in regulating transcription (GO:0006355, `http://amigo.geneontology.org/amigo/term/GO:0006355`) whereas less than 5% of sequences in group I have a similar annotation. The other significant metabolic processes involving these RRs are related to nucleotide and cyclic nucleotide biosynthetic processes, single organism metabolic processes (GO:0044710, `http://amigo.geneontology.org/amigo/term/GO:0044710`) and small molecular metabolic processes (GO:0044281, `http://amigo.geneontology.org/amigo/term/GO:0044281`.

Based on these annotations, we find that a large majority of multidomain HK regulated RRs have a predicted role in regulating transcription and with a direct DNA binding activity. Unlike multidomain HK mediated RRs, regular HK mediated RRs have nearly no predicted involvement in transcription and very little involvement in nucleotide metabolism. The only commonly predicted functional annotations for

regular HK mediated RRs is their involvement in ion binding and catalytic activity (see figure 4.6).

4.3 Summary of Findings

In this chapter, we examined 766 prokaryotic genomes and metagenomes and find that >70 % of genomes with sizes < 1 Mbp have no multidomain HKs. Using available data, we do not find any evidence to show that multidomain HKs are selectively preferred in smaller genomes over regular HKs (see figures 4.1B & 4.2), nor do we find a specific distribution of histidine kinase in specific environments. When we annotate sequences representing RRs regulated by different HK types, and their predicted process, function and location, we find that multidomain HKs are predominantly involved in nucleotide metabolism, regulate a number of catalytic, lyase and phosphorous-oxygen lyase activities (see figure 4.6) and directly regulate gene expression by binding to DNA (see figure 4.7).

4.4 Discussion & Conclusions

Following the theoretical findings discussed in chapters 2 and 3, we tried to find the biological significance of these findings and attempted to correlate them with different prokaryotic signalling architectures discussed before. By focussing on the input and response regulating members of these networks, we tried to correlate the theoretical signal processing properties of different signalling networks with the genomic occurrence of proteins found in two component signalling networks and the metabolic processes that the output from these networks control. With this, we tried to define the biological context for different prokaryotic signalling architectures.

To characterise the biological context of the input, we examined the distribution of different sensor proteins (HKs) across prokaryotic genomes and investigated whether smaller genomes have a preference for multidomain HKs (per Mbp) than larger genomes so as to leverage multistability in multidomain HK regulated networks and enable the cell to sense multiple signals and occupy different physiological states using fewer resources. With the data available on P2CS, we found that there wasn't sufficient evidence to show that there is a correlation between smaller genomes and the distribution of multidomain HKs in them.

To characterise the output regulated by these networks, we annotated sequences representing response regulators. We found that $> 60\%$ of multidomain HK regulated RRs have a predicted involvement in transcriptional regulation which indicates that signals which different multidomain HKs sense, have a direct impact on gene regulation when these regulators bind directly to the DNA. This indicates that genomes with multidomain HKs can directly leverage multistability and computational logic embedded in architectures containing multidomain HKs to control gene expression using shared architectures described in section 3.4. Such direct access to regulating gene expression via RRs make two component systems and phosphorelays ideal candidates for synthetic biology applications where engineered sensor and response regulators could be rewired [158, 159] to build synthetic circuits that implement novel ligand sensing and protein expression strategies.

5

Conclusions and Future Steps

In this study we discussed several of the most commonly occurring prokaryotic signalling architectures (figure 5.1). We built models representing commonly found four protein phosphorelay architectures to understand the significance of different biochemical reactions that occur in these networks, and the role that they play in prescribing different signal-response dynamics to the network. This exploration lead to the initial finding that some architectures have a capacity for tunability in their signal-response behaviour. Using steady state analytical solutions, we found specific mathematical relationships between reaction parameters that define when the necessary conditions for a system to exhibit either a sigmoidal or a hyperbolic response occur. Investigations like this describe how cells containing these commonly found phosphorelay architectures could employ mechanisms like regulating bifunctional kinase-phosphatase HK activity via ligand binding; modifying forward and reverse phosphotransfer reaction rates through mutations and regulating transcriptional feedback (which introduces new copies of proteins in the phosphorelay) to tune a signalling network's response behaviour from one response regime to another.

When we examine architectures which have multidomain HKs (figures 5.1B,C,E,F & G), we find that coupled double negative feedback loops occur in hybrid HK mediated phosphorelays (figure 5.1B) and embed bistability in the network. The occurrence of multistability in prokaryotic phosphotransfer mediated signalling networks was not previously known, and with proofs described in this study, we show

how signalling networks can embed unlimited multistability by sharing proteins, and extract specific relationships between the number of steady states and the number of signalling networks sharing the same protein (figure 5.1E, F and G). In addition to multistability, these networks with shared proteins can also compute logic functions like the AND, OR and ADDER functions.

As we examine findings from a simple two component signalling architecture (figure 5.1A) to a large network of networks with shared proteins (figure 5.1G), we see that the modularity of a simple two component architecture allows the creation of these larger networks, which give prokaryotes the ability to carry out a wide range of simple and complex signal processing, while using components much simpler than those seen in eukaryotic signalling.

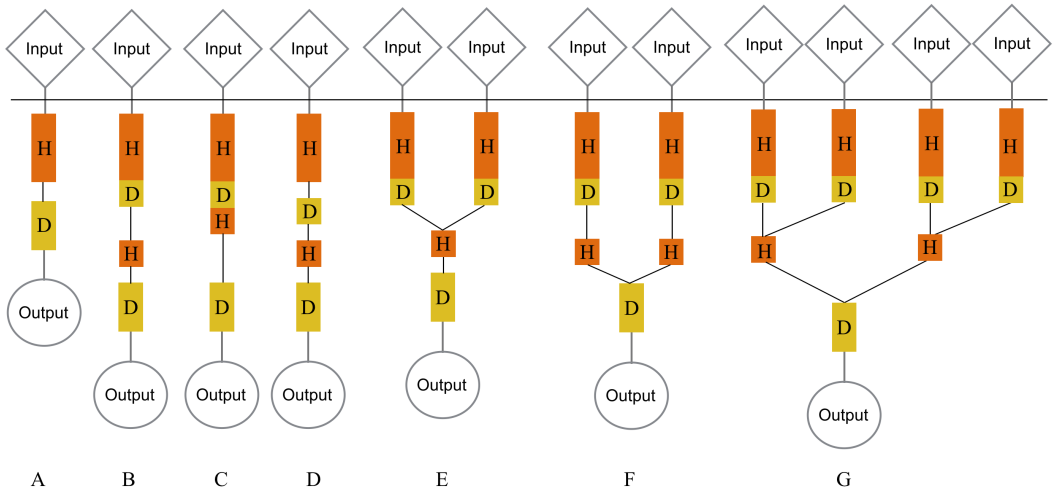


Figure 5.1: Shows key architectures which were modelled in this study. Simple TCS (A), hybrid HK mediated phosphorelay (B), unorthodox HK mediated phosphorelay (C), regular HK mediated phosphorelay (D), shared signalling networks with shared Hpt (E), shared RR(F) and shared Hpt and RR (G).

In addition to the insights that we have about the dynamic and computational properties that these signalling networks exhibit, their high modularity, specificity in signalling and a wide occurrence across hundreds of prokaryotic species make them an ideal target for applications in synthetic biology. So far, engineering of two-component systems have only concentrated on exploiting their signal sensing properties [160, 161] and not their signal processing capacities. In particular, the capacity for some of these architectures to implement multiple bistable switches that can be controlled at different signal levels or through a combination of signals, can inform the construction of novel synthetic logic gates at a protein level. With multistable phosphorelay networks, we could build controllable multistable systems

i.e. synthetic protein circuits and characterise them to extend the repertoire of tools available in synthetic biology.

List of publications, paper and poster presentations

Publications in Refereed Journals

1. Kothamachu VB, Feliu E, Cardelli L, Soyer OS (2015) Unlimited multistability and Boolean logic in microbial signalling. Journal of The Royal Society Interface 12(108):20150234.

2. Kothamachu VB, Feliu E, Wiuf C, Cardelli L, Soyer OS (2013) Phosphorelays Provide Tunable Signal Processing Capabilities for the Cell. PLoS Comput Biol 9(11):e1003322.

3. Amin M, et al. (2014) Phosphate Sink Containing Two-Component Signaling Systems as Tunable Threshold Devices. PLoS Comput Biol 10(10):e1003890.

Invited Talks

1. Invited talk titled "**Exploring Topological Properties of four-layered Phosphorelays**". BioComplexity Wednesday Meetings and Special Lectures, Neils Bohr Institute, Copenhagen. 13th March, 2013.

Conference Presentations

1. Contributed talk titled "**Sigmoidality, multistability and Boolean Logic in Two component-signalling architectures**". BioDynamics Workshop: Emergent Dynamics of Complex Biological Networks, 23-24 June 2014, University of Exeter

2. Poster Presentation titled "**Response Dynamics of Bacterial two component signalling systems**". Conference on Modelling in Microbiology at ESI, Edinburgh. 4-7 July, 2011.

3. Poster Presentation titled "**Response Dynamics of Bacterial two component signalling systems**". 6th Annual Summer School at Microsoft Research, Cambridge. 27th June-1st July, 2011

Appendices

A

APPENDIX

A.1 Parameters used to generate figure 2.1

Reaction rates	Hyperbolic Curve	Sigmoidal Curve
k_2 $(Ms)^{-1}$	9.96E+03	2.73E+04
k_3 $(Ms)^{-1}$	8.56E+04	1.66E+04
k_4 $(Ms)^{-1}$	8.78E+04	5.68E+04
k_{2r} $(Ms)^{-1}$	0	0
k_{3r} $(Ms)^{-1}$	1.68E+04	6.42E+04
k_{4r} $(Ms)^{-1}$	1.85E+04	9.71E+04
k_{h1} (s^{-1})	0.061	0.086
k_{h2} (s^{-1})	0.093	0.015
Hk_{tot} (M)	1.04E-04	1.77E-04
Rec_{tot} (M)	1.04E-04	1.77E-04
Hpt_{tot} (M)	1.04E-04	1.77E-04
RR_{tot} (M)	1.04E-04	1.77E-04

Table A.1: Parameters used to plot the curves in Figure 2.1.

	Total protein concentrations are equal		Total protein concentrations are unequal	
Parameters	Hyperbolic Regime	Sigmoidal Regime	Hyperbolic Regime	Sigmoidal Regime
k_2 $(Ms)^{-1}$	4.87E+04	5.15E+04	4.85E+04	5.21E+04
k_3 $(Ms)^{-1}$	6.00E+04	3.84E+04	5.98E+04	4.01E+04
k_4 $(Ms)^{-1}$	5.98E+04	3.88E+04	5.93E+04	3.90E+04
k_{2r} $(Ms)^{-1}$	0	0	0	0
k_{3r} $(Ms)^{-1}$	4.04E+04	6.05E+04	4.06E+04	6.06E+04
k_{4r} $(Ms)^{-1}$	4.14E+04	6.10E+04	4.17E+04	6.13E+04
k_{h1} (s^{-1})	0.057	0.052	0.059	0.054
k_{h2} (s^{-1})	0	0	0	0
Hk_{tot} (M)	1.03E-04	1.04E-04	9.49E-05	1.02E-04
REC_{tot} (M)	1.03E-04	1.04E-04	1.11E-04	1.09E-04
Hpt_{tot} (M)	1.03E-04	1.04E-04	9.75E-05	9.77E-05
RR_{tot} (M)	1.03E-04	1.04E-04	9.79E-05	9.71E-05
mean (k_2/k_{2r})	Inf	Inf	Inf	Inf
mean (k_3/k_{3r})	3.718	0.801	3.914	0.827
mean (k_4/k_{4r})	3.319	0.886	3.960	0.767

Table A.2: Mean parameter values calculated from parameter sampling runs implemented using numerical simulations for topology 14 (phosphorelays regulated by mono functional kinase).

	Total protein concentrations are equal		Total protein concentrations are unequal	
Parameters	Hyperbolic Regime	Sigmoidal Regime	Hyperbolic Regime	Sigmoidal Regime
k_2 $(Ms)^{-1}$	4.78E+04	6.83E+04	4.64E+04	6.76E+04
k_3 $(Ms)^{-1}$	5.27E+04	3.67E+04	5.23E+04	3.47E+04
k_4 $(Ms)^{-1}$	5.12E+04	3.52E+04	5.11E+04	3.51E+04
k_{2r} $(Ms)^{-1}$	5.47E+04	2.13E+04	5.69E+04	2.18E+04
k_{3r} $(Ms)^{-1}$	5.02E+04	6.19E+04	5.06E+04	6.39E+04
k_{4r} $(Ms)^{-1}$	4.96E+04	6.44E+04	5.05E+04	6.21E+04
k_{h1} (s^{-1})	0.055	0.051	0.054	0.054
k_{h2} (s^{-1})	0.000	0.000	0.000	0.000
HK_{tot} (M)	1.01E-04	1.05E-04	9.97E-05	1.06E-04
REC_{tot} (M)	1.01E-04	1.05E-04	1.07E-04	1.04E-04
Hpt_{tot} (M)	1.01E-04	1.05E-04	1.02E-04	9.84E-05
RR_{tot} (M)	1.01E-04	1.05E-04	9.98E-05	9.92E-05
mean (k_2/k_{2r})	2.453	11.147	1.757	15.092
mean (k_3/k_{3r})	2.577	0.708	2.650	0.626
mean (k_4/k_{4r})	2.218	0.696	2.842	0.778

Table A.3: Mean parameter values calculated from parameter sampling runs implemented using numerical simulations for topology 16 (phosphorelays regulated by mono functional kinase).

	Total protein concentrations are equal		Total protein concentrations are unequal	
Parameters	Hyperbolic Regime	Sigmoidal Regime	Hyperbolic Regime	Sigmoidal Regime
k_2 $(Ms)^{-1}$	4.85E+04	5.38E+04	4.97E+04	5.37E+04
k_3 $(Ms)^{-1}$	6.03E+04	3.91E+04	6.13E+04	3.77E+04
k_4 $(Ms)^{-1}$	5.98E+04	3.94E+04	5.85E+04	3.79E+04
k_{2r} $(Ms)^{-1}$	0.00E+00	0	0	0
k_{3r} $(Ms)^{-1}$	4.13E+04	6.08E+04	3.99E+04	6.06E+04
k_{4r} $(Ms)^{-1}$	3.92E+04	6.07E+04	4.08E+04	6.19E+04
k_{h1} (s^{-1})	0.049	0.052	0.050	0.051
k_{h2} (s^{-1})	0.051	0.049	0.052	0.050
HK_{tot} (M)	1.02E-04	1.09E-04	9.87E-05	1.11E-04
REC_{tot} (M)	1.015E-04	1.09E-04	9.84E-05	1.08E-04
Hpt_{tot} (M)	1.015E-04	1.09E-04	1.02E-04	1.02E-04
RR_{tot} (M)	1.02E-04	1.09E-04	1.08E-04	9.39E-05
mean (k_2/k_{2r})	Inf	Inf	Inf	Inf
mean (k_3/k_{3r})	5.55E+00	0.831	6.294	0.794
mean (k_4/k_{4r})	8.65E+00	0.842	6.250	0.824

Table A.4: Mean parameter values calculated from parameter sampling runs implemented using numerical simulations for topology 30 (phosphorelays regulated by mono functional kinase).

	Total protein concentrations are equal		Total protein concentrations are unequal	
Parameters	Hyperbolic Regime	Sigmoidal Regime	Hyperbolic Regime	Sigmoidal Regime
k_2 $(Ms)^{-1}$	4.95E+04	6.68E+04	5.08E+04	7.04E+04
k_3 $(Ms)^{-1}$	5.25E+04	3.56E+04	5.29E+04	3.14E+04
k_4 $(Ms)^{-1}$	5.34E+04	3.06E+04	5.19E+04	3.05E+04
k_{2r} $(Ms)^{-1}$	5.35E+04	1.85E+04	5.38E+04	2.07E+04
k_{3r} $(Ms)^{-1}$	4.81E+04	6.40E+04	4.87E+04	6.34E+04
k_{4r} $(Ms)^{-1}$	4.70E+04	6.03E+04	4.89E+04	5.87E+04
k_{h1} (s^{-1})	0.049	0.054	0.051	0.056
k_{h2} (s^{-1})	0.051	0.043	0.049	0.045
HK_{tot} (M)	1.01E-04	1.03E-04	9.99E-05	1.07E-04
REC_{tot} (M)	1.01E-04	1.03E-04	1.01E-04	1.07E-04
Hpt_{tot} (M)	1.01E-04	1.03E-04	9.98E-05	9.86E-05
RR_{tot} (M)	1.01E-04	1.03E-04	9.85E-05	9.99E-05
mean (k_2/k_{2r})	2.815	39.077	2.931	18.405
mean (k_3/k_{3r})	6.265	0.763	3.683	0.568
mean (k_4/k_{4r})	6.231	0.669	2.886	0.649

Table A.5: Mean parameter values calculated from parameter sampling runs implemented using numerical simulations for topology 32 (phosphorelays regulated by mono functional kinase).

	Total protein concentrations are equal		Total protein concentrations are unequal	
Parameters	Hyperbolic Regime	Sigmoidal Regime	Hyperbolic Regime	Sigmoidal Regime
k_2 $(Ms)^{-1}$	4.87E+04	5.15E+04	4.85E+04	5.21E+04
k_3 $(Ms)^{-1}$	6.00E+04	3.84E+04	5.98E+04	4.01E+04
k_4 $(Ms)^{-1}$	5.98E+04	3.88E+04	5.93E+04	3.90E+04
k_{2r} $(Ms)^{-1}$	0	0	0	0
k_{3r} $(Ms)^{-1}$	4.04E+04	6.05E+04	4.06E+04	6.06E+04
k_{4r} $(Ms)^{-1}$	4.14E+04	6.10E+04	4.17E+04	6.13E+04
k_{h1} (s^{-1})	0.057	0.052	0.059	0.054
k_{h2} (s^{-1})	0	0	0	0
HK_{tot} (M)	1.03E-04	1.04E-04	9.49E-05	1.02E-04
REC_{tot} (M)	1.03E-04	1.04E-04	1.11E-04	1.09E-04
Hpt_{tot} (M)	1.03E-04	1.04E-04	9.75E-05	9.77E-05
RR_{tot} (M)	1.03E-04	1.04E-04	9.79E-05	9.71E-05
mean (k_2/k_{2r})	Inf	Inf	Inf	Inf
mean (k_3/k_{3r})	3.718	0.801	3.914	0.827
mean (k_4/k_{4r})	3.319	0.886	3.960	0.767

Table A.6: Mean parameter values calculated from parameter sampling runs implemented using numerical simulations for topology 14 (phosphorelays regulated by bi-functional kinase).

	Total protein concentrations are equal		Total protein concentrations are unequal	
Parameters	Hyperbolic Regime	Sigmoidal Regime	Hyperbolic Regime	Sigmoidal Regime
k_2 $(Ms)^{-1}$	4.78E+04	6.83E+04	4.64E+04	6.76E+04
k_3 $(Ms)^{-1}$	5.27E+04	3.67E+04	5.23E+04	3.47E+04
k_4 $(Ms)^{-1}$	5.12E+04	3.52E+04	5.11E+04	3.51E+04
k_{2r} $(Ms)^{-1}$	5.47E+04	2.13E+04	5.69E+04	2.18E+04
k_{3r} $(Ms)^{-1}$	5.02E+04	6.19E+04	5.06E+04	6.39E+04
k_{4r} $(Ms)^{-1}$	4.96E+04	6.44E+04	5.05E+04	6.21E+04
k_{h1} (s^{-1})	0.055	0.051	0.054	0.054
k_{h2} (s^{-1})	0.000	0.000	0.000	0.000
HK_{tot} (M)	1.01E-04	1.05E-04	9.97E-05	1.06E-04
REC_{tot} (M)	1.01E-04	1.05E-04	1.07E-04	1.04E-04
Hpt_{tot} (M)	1.01E-04	1.05E-04	1.02E-04	9.84E-05
RR_{tot} (M)	1.01E-04	1.05E-04	9.98E-05	9.92E-05
mean (k_2/k_{2r})	2.453	11.147	1.757	15.092
mean (k_3/k_{3r})	2.577	0.708	2.650	0.626
mean (k_4/k_{4r})	2.218	0.696	2.842	0.778

Table A.7: Mean parameter values calculated from parameter sampling runs implemented using numerical simulations for topology 16 (phosphorelays regulated by bi-functional kinase).

	Total protein concentrations are equal		Total protein concentrations are unequal	
Parameters	Hyperbolic Regime	Sigmoidal Regime	Hyperbolic Regime	Sigmoidal Regime
k_2 $(Ms)^{-1}$	4.85E+04	5.38E+04	4.97E+04	5.37E+04
k_3 $(Ms)^{-1}$	6.03E+04	3.91E+04	6.13E+04	3.77E+04
k_4 $(Ms)^{-1}$	5.98E+04	3.94E+04	5.85E+04	3.79E+04
k_{2r} $(Ms)^{-1}$	0.00E+00	0	0	0
k_{3r} $(Ms)^{-1}$	4.13E+04	6.08E+04	3.99E+04	6.06E+04
k_{4r} $(Ms)^{-1}$	3.92E+04	6.07E+04	4.08E+04	6.19E+04
k_{h1} (s^{-1})	0.049	0.052	0.050	0.051
k_{h2} (s^{-1})	0.051	0.049	0.052	0.050
HK_{tot} (M)	1.02E-04	1.09E-04	9.87E-05	1.11E-04
REC_{tot} (M)	1.015E-04	1.09E-04	9.84E-05	1.08E-04
Hpt_{tot} (M)	1.015E-04	1.09E-04	1.02E-04	1.02E-04
RR_{tot} (M)	1.02E-04	1.09E-04	1.08E-04	9.39E-05
mean (k_2/k_{2r})	Inf	Inf	Inf	Inf
mean (k_3/k_{3r})	5.55E+00	0.831	6.294	0.794
mean (k_4/k_{4r})	8.65E+00	0.842	6.250	0.824

Table A.8: Mean parameter values calculated from parameter sampling runs implemented using numerical simulations for topology 30 (phosphorelays regulated by bi-functional kinase).

	Total protein concentrations are equal		Total protein concentrations are unequal	
Parameters	Hyperbolic Regime	Sigmoidal Regime	Hyperbolic Regime	Sigmoidal Regime
k_2 $(Ms)^{-1}$	4.95E+04	6.68E+04	5.08E+04	7.04E+04
k_3 $(Ms)^{-1}$	5.25E+04	3.56E+04	5.29E+04	3.14E+04
k_4 $(Ms)^{-1}$	5.34E+04	3.06E+04	5.19E+04	3.05E+04
k_{2r} $(Ms)^{-1}$	5.35E+04	1.85E+04	5.38E+04	2.07E+04
k_{3r} $(Ms)^{-1}$	4.81E+04	6.40E+04	4.87E+04	6.34E+04
k_{4r} $(Ms)^{-1}$	4.70E+04	6.03E+04	4.89E+04	5.87E+04
k_{h1} (s^{-1})	0.049	0.054	0.051	0.056
k_{h2} (s^{-1})	0.051	0.043	0.049	0.045
HK_{tot} (M)	1.01E-04	1.03E-04	9.99E-05	1.07E-04
REC_{tot}	1.01E-04	1.03E-04	1.01E-04	1.07E-04
Hpt_{tot} (M)	1.01E-04	1.03E-04	9.98E-05	9.86E-05
RR_{tot}	1.01E-04	1.03E-04	9.85E-05	9.99E-05
mean (k_2/k_{2r})	2.815	39.077	2.931	18.405
mean (k_3/k_{3r})	6.265	0.763	3.683	0.568
mean (k_4/k_{4r})	6.231	0.669	2.886	0.649

Table A.9: Mean parameter values calculated from parameter sampling runs implemented using numerical simulations for topology 32 (phosphorelays regulated by bi-functional kinase).

A.2 Parameters used to generate figure 2.6

Parameters used to generate panels A-F in figure 2.6 are listed below. The values are given in the order k_2, k_3, k_4, k_{2r}, k_{3r},k_{4r}, k_{h1}, k_{h2}, k_5, k_{5r}, k_6, HK_{tot}, REC_{tot}, Hpt_{tot}, RR_{tot}:

Panel A) (9343, 30201, 35826, 0, 7192, 99251, 0.0302, 0.00234, 1000, 0.012, 3.5, 1.2748e-04, 1.5755e-04, 1.3634e-04, 1.2516e-04)

Panel B) (186860, 90605, 35827, 0, 35963, 49626, 0.0302, 0.0023, 1000, 0.012, 3.5, 1.2748e-04, 1.5755e-04, 1.363e-04, 1.251e-04)

Panel C) (5, 0.1, 0.01, 0, 10, 0.10, 10, 0.001,-,-,-, 5, 1, 1, 10)

Panel D) (500, 0.1, 0.01, 0, 1, 0.10, 0.1, 0.001,-,-,-, 1, 0.005, 100000, 100)

Panel E) (5000, 0.1, 1, 0, 1, 0, 10, 100, 0, -,-,-,1, 10, 2000, 1)

Panel F) (5000, 0.1, 0.01, 0, 1, 0.001, 10, 1,-,-,-, 1, 0.1, 1, 30).

B
APPENDIX

B.1 Proof of the claims made in section 2.2.1

B.1.1 Constant signal-response curves and zero solutions

- $k_{h1} = k_{h2} = 0$: From (46) we have that $x_1 = 0$ (since $k_s \neq 0$) and thus $x_2 = \overline{H}$ at steady state. From (47) it follows that $x_3 = 0$ (since $k_2 \neq 0$) and thus $x_4 = \overline{C}$. From (48) we have that $x_5 = 0$ (since $k_3 \neq 0$) and so $x_6 = \overline{T}$ and finally from (49) it follows that $x_7 = 0$ (since $k_4 \neq 0$) and $x_8 = \overline{R}$. This fact is independent of the value of all other parameters.

- $k_{h2} = k_{3r} = 0$: From (48) either $x_4 = 0$ or $x_5 = 0$. If $x_4 = 0$ we have that $x_3 \neq 0$. From (46), $x_1 = 0$ and from (47) we have $x_2 = 0$ which is a contradiction. Therefore $x_5 = 0$ and consequently, $x_6 = \overline{T} \neq 0$ and $x_7 = 0$. Thus $x_8 = \overline{R}$ and $x_1, \ldots, x_4$ fulfill (46), (47) and the equations for the total amounts. Additionally, the signal-response curve for Hpt is also constant (that is, at steady state $x_6 = \overline{T}$), but the curves corresponding to HK and REC are not constant.

- $k_{h2} = k_{4r} = 0$: From (49) either $x_6 = 0$ or $x_7 = 0$. If $x_6 = 0$ we have that $x_5 \neq 0$. From (48) $x_4 = 0$ and from (46) we have $x_1 = 0$. By (47) we have $x_2 x_3 = 0$ which is a contradiction. Therefore $x_7 = 0$, and then $x_8 = \overline{R} \neq 0$. $x_1, \ldots, x_6$ fulfill (46)-(48) and the equations for the total amounts. In this case

the signal-response curves for HK, REC and Hpt are not constant.

Assume that none of the cases above hold, that is, that either $k_{h2} \neq 0$ or $k_{h2} = 0$ and $k_{h1}k_{3r}k_{4r} \neq 0$. Assume that all total amounts are positive. We show that in this case any non-negative solution to the steady-state equations is positive, that is, all concentrations are non-zero. As a consequence, the signal-response curve cannot be constant (equal to $\overline{R}$; because this would imply $x_7 = 0$).

- If $x_1 = 0$, then by (46) it must be that $k_{h1}x_4 = k_{h2}x_8 = 0$ and $x_2 \neq 0$. From (47) it follows that $x_3 = 0$ and hence $x_4 \neq 0$ (because $\overline{C} > 0$) and thus $k_{h1} = 0$. In this case $k_{h2} \neq 0$ and hence $x_8 = 0$. From (48) we have $x_5 = 0$ and hence $x_6 \neq 0$. From (49) we see that $x_7 = 0$ contradicting $\overline{R} > 0$.

- If $x_2 = 0$ or $x_3 = 0$ then from (47) we have $k_{h1}x_4 = k_{h2}x_8 = k_{2r}x_1x_4 = 0$. It follows from (46) that $x_1 = 0$ and we reach a contradiction with the item above.

- If $x_4 = 0$ or $x_5 = 0$ then from (48) we have $k_{3r}x_3x_6 = k_{h2}x_8 = 0$. If $x_4 = 0$ using (46) we have that $k_s x_1 = 0$ and hence $x_1 = 0$, which is a contradiction. If $x_5 = 0$ then from (49) we have $k_4x_6x_7 = 0$. Since $x_6 \neq 0$ (because $\overline{T} > 0$) and $k_4 \neq 0$ by hypothesis, we have $x_7 = 0$. As a consequence $x_8 \neq 0$. Hence $k_{h2} = 0$ and by hypothesis $k_{h1}k_{3r}k_{4r} \neq 0$. If $k_{3r} \neq 0$ we have $x_3 = 0$ contradicting the item above.

- If $x_6 = 0$ or $x_7 = 0$ then $k_{h2}x_8 = k_{4r}x_5x_8 = 0$. Since we showed that $x_5 \neq 0$, and $k_{h2}k_{4r} \neq 0$ by hypothesis, it follows that $x_8 = 0$. If $x_7 = 0$ we reach a contradiction. If $x_6 = 0$ then using (48) we have $x_4x_5 = 0$ which contradicts the item above.

- If $x_8 = 0$ then $x_7 \neq 0$ and hence by (49) $x_6 = 0$ which contradicts the item above.

B.1.2 Hyperbolic shape when phosphorelay rates are large

We prove here that if (62) holds, then the second derivative of φ at any value of k_s is negative.
We compute the second derivative of x_4 with respect to x_8 and find that its sign equals the sign of

$$(k_3 - k_{3r})(k_4k_{h2}\overline{R} + k_{3r}k_{h2}\overline{C}) + k_{3r}(k_3k_4 - k_{3r}k_{4r})\overline{C} \cdot \overline{T}.$$

Consider now

$$x_2 = \frac{k_{2r}\overline{H}x_4 + k_{h1}x_4 + k_{h2}x_8}{k_2x_3 + k_{2r}x_4} = \frac{(k_{2r}\overline{H} + k_{h1})x_4 + k_{h2}x_8}{(k_{2r} - k_2)x_4 + k_2\overline{C}}.$$

Let $\beta = (k_{2r}\overline{H} + k_{h1})x_4 + k_{h2}x_8$ be the numerator of x_2 and $\gamma = (k_{2r} - k_2)x_4 + k_2\overline{C}$ be the denominator of x_2. Both terms are positive. We take the second derivative of x_2 with respect to x_8 and obtain:

$$\begin{aligned} x_2'' &= \frac{(\beta''\gamma - \beta\gamma'')\gamma - 2\gamma'(\beta'\gamma - \beta\gamma')}{\gamma^3} \\ &= \frac{((k_{2r}\overline{H} + k_{h1})\gamma - (k_{2r} - k_2)\beta)\gamma x_4'' - 2(k_{2r} - k_2)x_4'(\beta'\gamma - \beta\gamma')}{\gamma^3}. \end{aligned}$$

Let

$$\begin{aligned} A &= ((k_{2r}\overline{H} + k_{h1})\gamma - (k_{2r} - k_2)\beta)\gamma x_4'' = (k_{2r}(\overline{H}\gamma - \beta) + k_{h1}\gamma + k_2\beta)\gamma x_4'', \\ B &= -2(k_{2r} - k_2)x_4'(\beta'\gamma - \beta\gamma') \end{aligned}$$

such that $x_2'' = (A + B)/\gamma^3$. The denominator is positive. Therefore, the sign of x_2'' is determined by the sign of $A + B$. Since x_2, x_4 increase in x_8 we have that $\beta'\gamma - \beta\gamma' > 0$ and $x_4' > 0$. Therefore, the sign of B equals the sign of $k_2 - k_{2r}$.
The term $\overline{H}\gamma - \beta = -k_2x_4\overline{H} + k_2\overline{C} \cdot \overline{H} - k_{h1}x_4 - k_{h2}x_8$ is positive in I because it agrees with the numerator of x_1. It follows that the sign of A equals the sign of x_4''. If the signs of A and B agree, then x_2'' has a constant sign over I.
Consider now the inverse of the signal-response curve:

$$k_s = f(x_8) := \frac{k_{h1}x_4 + k_{h2}x_8}{x_1}$$

and let $\delta = k_{h1}x_4 + k_{h2}x_8$. The second derivative of f with respect to x_8 is

$$f'' = \frac{x_1(k_{h1}x_4''x_1 - \delta x_1'') \ - 2x_1'(\delta' x_1 - \delta x_1')}{x_1^3}.$$

The term $-2x_1'(\delta' x_1 - \delta x_1')$ is positive because $x_1' < 0$ and $(\delta' x_1 - \delta x_1') > 0$ (it is the numerator of the derivative of f). If $(k_{h1}x_4''x_1 - \delta x_1'') > 0$, then the signal-response curve is hyperbolic (because the sign of the second derivative of φ is minus the sign of the second derivative of f). In particular, this is the case if $x_4'' > 0$ and $x_1'' < 0$. For $x_1'' < 0$ we require $x_2'' > 0$.
Therefore, if $x_2'', x_4'' > 0$, then the signal-response curve is hyperbolic. Using the computations above, we conclude that if

$$k_2 - k_{2r} > 0, \qquad \text{and} \qquad (k_3 - k_{3r})(k_4k_{h2}\overline{R} + k_{3r}k_{h2}\overline{C}) + k_{3r}(k_3k_4 - k_{3r}k_{4r})\overline{CT} > 0,$$

then the curve is hyperbolic. These two inequalities are in particular fulfilled if

$$k_2 > k_{2r}, \qquad k_3 > k_{3r}, \qquad k_4 > k_{4r},$$

that is, if the forward phosphorelay rate constants are larger than their reverse counterparts.

C

APPENDIX

C.1 Analytical solution for phosphorelays with bifunctional HK (section 2.2.3)

We consider the case in which the kinase HK is bifunctional, that is, HK acts as a phosphatase for REC.

C.1.1 Reactions, equations and steady states

Reactions. The minimal set of reactions that the system has, consists of the autophosphorylation reaction on HK and the forward phosphotransfer reactions

$$\mathrm{HK} \xrightarrow{k_s} \mathrm{HKp} \qquad \mathrm{HKp} + \mathrm{REC} \xrightarrow{k_2} \mathrm{HK} + \mathrm{RECp}$$

$$\mathrm{RECp} + \mathrm{Hpt} \xrightarrow{k_3} \mathrm{REC} + \mathrm{Hptp} \qquad \mathrm{Hptp} + \mathrm{RR} \xrightarrow{k_4} \mathrm{Hpt} + \mathrm{RRp},$$

together with the reaction for phosphatase activity of the histidine kinase HK:

$$\mathrm{HK} + \mathrm{RECp} \underset{\mathrm{k_5 r}}{\overset{\mathrm{k_5}}{\rightleftharpoons}} \mathrm{HK} - \mathrm{RECp} \xrightarrow{k_6} \mathrm{HK} + \mathrm{REC}.$$

We study the extensions of this minimal reaction mechanism obtained by adding reverse phospho-transfer reactions involving the HK, REC, Hpt and the RR, as well

as hydrolysis reactions for RECp and RRp:

$$\begin{aligned} &\mathrm{HK} + \mathrm{RECp} \xrightarrow{k_{2r}} \mathrm{HKp} + \mathrm{REC} \qquad \mathrm{RECp} \xrightarrow{k_{h1}} \mathrm{REC} \\ &\mathrm{REC} + \mathrm{Hptp} \xrightarrow{k_{3r}} \mathrm{RECp} + \mathrm{Hpt} \qquad \mathrm{RRp} \xrightarrow{k_{h2}} \mathrm{RR} \\ &\mathrm{Hpt} + \mathrm{RRp} \xrightarrow{k_{4r}} \mathrm{Hptp} + \mathrm{RR}. \end{aligned}$$

By setting some of the rate constants k_{*r} or k_{h*} to zero, we obtain a total of 32 different topologies involving different combinations of reverse phosphorelay and hydrolysis reactions, similarly to the situation where the HK kinase is monofunctional (table 2.1). The minimal set of reactions is always part of the system, meaning that $k_2, k_3, k_4, k_s \neq 0$ for all topologies.

Ordinary differential equations. We model the protein concentrations in the system using ordinary differential equations (ODEs). To simplify the notation, we define:

$$\begin{aligned} &x_1 = [\mathrm{HK}], \quad x_2 = [\mathrm{HKp}], \quad x_3 = [\mathrm{REC}], \quad x_4 = [\mathrm{RECp}], \\ &x_5 = [\mathrm{Hpt}], \quad x_6 = [\mathrm{Hptp}], \quad x_7 = [\mathrm{RR}], \quad x_8 = [\mathrm{RRp}], \quad x_9 = [\mathrm{HKRECp}]. \end{aligned}$$

The dynamics of the concentrations in time is modeled with a system of ODEs:

$$\begin{aligned} \dot{x}_1 &= -k_s x_1 - k_{2r} x_1 x_4 + k_2 x_2 x_3 - k_5 x_1 x_4 + k_{5r} x_9 + k_6 x_9 \\ \dot{x}_2 &= k_s x_1 + k_{2r} x_1 x_4 - k_2 x_2 x_3 \\ \dot{x}_3 &= -k_2 x_2 x_3 + k h_1 x_4 + k_{2r} x_1 x_4 + k_3 x_4 x_5 - k_{3r} x_3 x_6 + k_6 x_9 \\ \dot{x}_4 &= k_2 x_2 x_3 - k_{h1} x_4 - k_{2r} x_1 x_4 - k_3 x_4 x_5 + k_{3r} x_3 x_6 - k_5 x_1 x_4 + k_{5r} x_9 \\ \dot{x}_5 &= -k_3 x_4 x_5 + k_{3r} x_3 x_6 + k_4 x_6 x_7 - k_{4r} x_5 x_8 \\ \dot{x}_6 &= k_3 x_4 x_5 - k_{3r} x_3 x_6 - k_4 x_6 x_7 + k_{4r} x_5 x_8 \\ \dot{x}_7 &= -k_4 x_6 x_7 + k_{h2} x_8 + k_{4r} x_5 x_8 \\ \dot{x}_8 &= k_4 x_6 x_7 - k_{h2} x_8 - k_{4r} x_5 x_8 \\ \dot{x}_9 &= k_5 x_1 x_4 - k_{5r} x_9 - k_6 x_9. \end{aligned}$$

This system has four conserved amounts:

$$\overline{H} = x_1 + x_2 + x_9, \quad \overline{C} = x_3 + x_4 + x_9, \quad \overline{T} = x_5 + x_6, \quad \overline{R} = x_7 + x_8.$$

It is assumed that total amounts are positive.

Steady-state equations. We proceed as in the monofunctional case and conclude that the steady states of the system are given as the solutions to the following system of equations:

$$\overline{H} = x_1 + x_2 + x_9 \tag{186}$$

$$\overline{C} = x_3 + x_4 + x_9 \tag{187}$$

$$\overline{T} = x_5 + x_6 \tag{188}$$

$$\overline{R} = x_7 + x_8 \tag{189}$$

$$0 = k_s x_1 + k_{2r} x_1 x_4 - k_2 x_2 x_3 \tag{190}$$

$$0 = k_2 x_2 x_3 - k_{h1} x_4 - k_{2r} x_1 x_4 - k_3 x_4 x_5 + k_{3r} x_3 x_6 - k_5 x_1 x_4 + k_{5r} x_9 \tag{191}$$

$$0 = k_3 x_4 x_5 - k_{3r} x_3 x_6 - k_4 x_6 x_7 + k_{4r} x_5 x_8 \tag{192}$$

$$0 = k_4 x_6 x_7 - k_{h2} x_8 - k_{4r} x_5 x_8 \tag{193}$$

$$0 = k_5 x_1 x_4 - k_{5r} x_9 - k_6 x_9. \tag{194}$$

Rearrangement of the steady-state equations. We change equations (190)-(194) by linear combinations of them. Specifically, we replace:

- (192) by (192)+(193),
- (191) by (191)+(192)+(193)+(194),
- (190) by (190)+(191)+(192)+(193)+(194),

and leave (193) and (194) as they are. This results in the following equivalent system of equations:

$$\overline{H} = x_1 + x_2 + x_9 \tag{195}$$

$$\overline{C} = x_3 + x_4 + x_9 \tag{196}$$

$$\overline{T} = x_5 + x_6 \tag{197}$$

$$\overline{R} = x_7 + x_8 \tag{198}$$

$$0 = k_s x_1 - k_{h1} x_4 - k_{h2} x_8 - k_6 x_9 \tag{199}$$

$$0 = k_2 x_2 x_3 - k_{h1} x_4 - k_{2r} x_1 x_4 - k_{h2} x_8 - k_6 x_9 \tag{200}$$

$$0 = k_3 x_4 x_5 - k_{3r} x_3 x_6 - k_{h2} x_8 \tag{201}$$

$$0 = k_4 x_6 x_7 - k_{h2} x_8 - k_{4r} x_5 x_8 \tag{202}$$

$$0 = k_5 x_1 x_4 - k_{5r} x_9 - k_6 x_9. \tag{203}$$

Zero concentrations. We assume that $k_5, k_{5r}, k_6 \neq 0$, that is, the kinase is bifunctional and acts as a phosphatase for the dephosphorylation of the receiver protein. Additionally, we are assuming that $k_s, k_2, k_3, k_4 \neq 0$ and that all total amounts are positive. In this scenario, for topologies 9-13 and 15, the signal-response curve is constant and equals $x_8 = \overline{R}$. All the other topologies have non-constant signal-response curves and all concentrations at steady state are non-zero (see Appendix C.2.1 for proofs).

Topologies 17-24 ($k_{h1} = k_{h2} = 0$) exhibit signal-response curves defined piecewise and are treated differently. Zero steady-state values occur (see Appendix C.2.3 for proofs).

C.1.2 Steady-state relations and signal-response curves

We proceed to find an expression for the signal-response curve. As in the previous case, the *signal* is taken to be the value of the rate constant k_s, and the *response* is the steady-state value of phosphorylated response regulator (x_8) corresponding to k_s (with all the other rate constants and total amounts fixed). Contrary to the previous system, this system with the bifunctional kinase does not allow for an explicit analytical relation. Instead, we infer the existence of an analytical function relating k_s and x_8 and derive properties of this function. We start by establishing relations between each concentration x_i and x_8 at steady state.

Steady-state relations. We study here the steady-state solutions that do not have vanishing concentrations. We let

$$k_y = \frac{k_5}{k_{5r} + k_6} \tag{204}$$

be the reciprocal of the Michaelis-Menten constant of HK. We express the concentrations of $x_1, \ldots, x_7$ at steady state in terms of x_8 and x_9, independently of k_s. In addition, we find a relation between x_8 and x_9 at steady state.

Expression	Behavior as function of x_8 and x_9
$x_7 = \overline{R} - x_8$	x_7 decreases in x_8
$x_6 = \frac{x_8(k_{4r}\overline{T}+k_{h2})}{k_4 x_7 + k_{4r} x_8}$	x_6 increases in x_8
$x_5 = \frac{k_4\overline{T}x_7 - k_{h2}x_8}{k_4 x_7 + k_{4r} x_8}$	x_5 decreases in x_8
$x_4 = \frac{k_{h2}x_8 + k_{3r}x_6(C-x_9)}{k_3 x_5 + k_{3r} x_6}$	x_4 increases in x_8 and decreases in x_9
$x_3 = \frac{k_3 x_5 (C-x_9) - k_{h2}x_8}{k_3 x_5 + k_{3r} x_6}$	x_3 decreases in x_8 and in x_9
$x_2 = \frac{k_{2r}x_4(\overline{H}-x_9) + k_6 x_9 + k_{h1}x_4 + k_{h2}x_8}{k_2 x_3 + k_{2r} x_4}$	x_2 increases in x_8
$x_1 = \frac{k_2 x_3(\overline{H}-x_9) - k_6 x_9 - k_{h1}x_4 - k_{h2}x_8}{k_2 x_3 + k_{2r} x_4}$	x_1 decreases in x_8 and in x_9
$x_9 = g(x_8)$	

(205)

The first seven rows of (205) give an iterative way to find the steady-state values of the concentrations of $x_1, \ldots, x_7$ once the values of x_8 and x_9 are known. The last entry gives the relation between x_9 and x_8. See Appendix C.2.2 for a proof.

The steady-state values are all positive if and only if x_8 is in the interval $(0, \alpha)$,

where α is the first positive root of the degree-2 polynomial q_2 given in (57) (that is, the same as for the monofunctional case).

In (205), the variables $x_1, \ldots, x_7$ are expressed as functions of x_8, x_9. The variable x_9 cannot explicitly be written as a function of x_8. The function g is known to exist, but we do not have an analytical expression for it. However, there is a procedure to obtain the steady-state value of x_9 corresponding to a given value of x_8. For each fixed x_8 strictly between 0 and α, x_9 is the first positive root of the following polynomial $G(x_8, x_9)$:

$$G(x_8, x_9) = c_0(x_8) + c_1(x_8)x_9 + c_2(x_8)x_9^2 + c_3(x_8)x_9^3,$$

where if we denote

$$z_1 := k_{3r}(k_{4r}T + k_{h2}), \quad z_2 := k_2k_3(k_4T + k_{h2}), \quad z_3 := k_{4r} - k_4, \quad z_4 := z_1 - k_3(k_4T + k_{h2}),$$

then the coefficients $c_i(x_8)$ $1 \leq i \leq 3$ are:

$$\begin{aligned}
c_0(x_8) =& k_y x_8(k_{h2}(k_4\overline{R} + z_3x_8) + z_1\overline{C})\Big(k_{h2}(z_3(k_{h1} + k_2\overline{H}) + z_4)x_8^2 \\
&+ (\overline{C}(k_{h1}z_1 + z_2\overline{H}) + k_{h2}k_4\overline{R}(k_2\overline{H} + k_3\overline{T} + k_{h1}))x_8 - k_2k_3k_4\overline{CHTR}\Big) \\
c_1(x_8) =& \Big((k_{2r}z_1 - z_2)\overline{C}x_8 - k_{h2}(k_2 - k_{2r})(k_4\overline{R} + z_3x_8)x_8 + k_2k_3k_4\overline{CTR}\Big)(k_3k_4\overline{TR} + z_4x_8) \\
&+ \Big((k_2k_3k_4\overline{TR} - x_8z_2)(k_{h2}(x_8z_3 + k_4\overline{R})(\overline{H} + \overline{C}) + z_1(2\overline{H} + \overline{C})\overline{C}) \\
&- k_{h2}k_2z_1(x_8z_3 + k_4\overline{R})(\overline{H} + \overline{C})x_8 - k_2k_{h2}^2x_8(x_8^2z_3^2 + k_4^2\overline{R}^2) \\
&- 2k_{h2}z_3(k_2k_4k_{h2}\overline{R} + z_1k_{h1})x_8^2 + k_{h2}x_8(k_6z_3 - z_1)(k_3k_4\overline{TR} + z_4x_8) \\
&+ (k_6z_4x_8 - 2z_1k_{h1}x_8 + k_3k_4k_6\overline{TR})(z_1\overline{C} + k_{h2}k_4\overline{R})\Big)x_8k_y \\
c_2(x_8) =& \Big(k_{h2}z_3(z_2 + k_2z_1)x_8^2 + (k_{h1}z_1^2 + (2z_2\overline{C} + k_2k_{h2}k_4\overline{R} - k_6z_4 + z_2\overline{H})z_1 \\
&- k_4k_{h2}\overline{R}(k_2k_3z_3\overline{T} - z_2))x_8 - k_3k_4\overline{RT}(k_2k_4k_{h2}\overline{R} + z_1(k_6 + 2\overline{C}k_2 + k_2\overline{H}))\Big)x_8k_y \\
&- (k_3k_4\overline{TR} + z_4x_8)(k_2k_3k_4\overline{TR} + (k_{2r}z_1 - z_2)x_8) \\
c_3(x_8) =& z_1(k_2k_3k_4\overline{TR} - z_2x_8)k_yx_8
\end{aligned}$$

Signal-response expression. Using the remaining steady-state equation, (199), we express k_s in terms of x_8:

$$k_s = f_b(x_8) = \frac{k_{h1}x_4 + k_{h2}x_8 + k_6x_9}{x_1}, \tag{206}$$

where x_1, x_4, x_9 are given in terms of x_8 as well (see (205) above) and $x_1 \neq 0$ at

steady state. If x_8 approaches α, then x_1 tends to zero and k_s tends to infinity. It follows that α is precisely the maximal response of x_8.
The function f_b is continuous and differentiable in $[0, \alpha)$ and is strictly increasing (see Appendix C.2.3). It admits an inverse

$$\varphi_b = f_b^{-1},$$

which is the signal-response curve. The signal-response curve is increasing, continuous and differentiable in $[0, +\infty)$. When k_s tends to infinity then the response x_8 tends to α.
If $k_{h1} = k_{h2} = 0$ then φ_b is defined by f_b^{-1} if $k_s \in [0, k_6 k_y \overline{C}]$ and $\varphi_b = \overline{R}$ for $k_s > k_6 k_y \overline{C}$.

Practical considerations. In order to plot the signal-response curve we use the following procedure:

(i) Compute α (the first positive root of $q_2(x_8)$ in x_8) and choose a grid of values for x_8, strictly between 0 and α.

(ii) For each value of x_8, find the first positive root of $G(x_8, x_9)$ as a function of x_9, that is, for each value of x_8 we find a value of x_9.

(iii) Compute x_1, x_4 using (205) and the pair of values (x_8, x_9).

(iv) Compute k_s using (206) in terms of x_1, x_4, x_8, x_9.

In this way, points on the signal-response curve $(f_b(x_8), x_8)$ are generated. Because of the relationship between f_b and φ_b, the points give a plot of the function φ_b.

C.1.3 Hyperbolic and sigmoidal signal-response curves

We apply the same indicator to classify a curve as sigmoidal or hyperbolic as in the previous case. That is, we calculate the sign of the second derivative of the signal-response curve at zero and classify the curve accordingly. We have computed $\varphi_b''(0)$ using the method introduced in section 2.3.1.

If $\varphi_b''(0) > 0$ then we classify the signal-response curve as sigmoidal, and if $\varphi_b''(0) < 0$ then we classify the signal-response curve as sigmoidal (section 1.2.5).

The sign of the second derivative of φ_b at 0 agrees with the sign of

$$S_b = S + H(\alpha_2 H + \alpha_3)k_y + \alpha_1 H k_y^2 \tag{207}$$

where $k_y = \frac{k_5}{k_{5r}+k_6}$ and

$$\begin{aligned}
\alpha_1 =& C(k_{4r}T + k_{h2})H^2 k_2 k_6 k_{3r}(k_{3r}C(k_{4r}T + k_{h2}) + k_4 k_{h2} R) \\
\alpha_2 =& -C(k_{4r}T + k_{h2})k_2(Ck_3k_6k_{3r}(k_4T + k_{h2}) - k_{3r}^2(k_{h1}C + k_6)(k_{4r}T + k_{h2}) \\
&+ k_4 k_{h2} R(k_6(k_3 - k_{3r}) - k_{h1}k_{3r})) \\
\alpha_3 =& -(k_{3r}C(k_{4r}T + k_{h2}) + k_4 k_{h2} R)(k_{3r}k_{h1}(k_{4r}CT + k_{h2}) + k_4 k_{h2} R(k_3 T + k_{h1}))(k_2 C + k_6)
\end{aligned}$$

The term α_1 is always positive and the term α_3 is always negative. The independent term (obtained by setting $k_y = 0$) is identical to the term given in the monofunctional case (eq (60)). We have that if $k_{3r} = 0$, then $S_b < 0$ and the function φ_b is hyperbolic. However, when $k_{h1} = 0$ the system can show sigmoidality (because $\alpha_1 \neq 0$).

Observe that the leading coefficient of the term S_b in (207) seen as a polynomial of degree 2 in k_y is positive. Therefore, by increasing k_y enough, S_b becomes positive and the curve sigmoidal. Recall that k_y is the reciprocal of the Michaelis-Menten constant of the enzyme HK for its dephosphorylation activity. Therefore, increasing k_y corresponds to making the enzyme mediated dephosphorylation of RECp more efficient.

C.2 Proof of the claims: bifunctional case

This section provides a sketch of the proofs for the claims made earlier in section C.1.

C.2.1 Zero concentrations

We start by checking that the combinations $k_{h2} = 0$ and either $k_{3r} = 0$ or $k_{4r} = 0$ provide constant signal-response curves. Assume that $k_{h2} = 0$ and $x_4 = 0$ at steady state. Then by (203) $x_9 = 0$. Consequently from (199) we have $x_1 = 0$ and hence $x_2 \neq 0$ (195). Similarly from (196) we have that $x_3 \neq 0$. But then (200) cannot hold. Therefore, if $k_{h2} = 0$, $x_4 \neq 0$ at steady state.

- $k_{h2} = k_{3r} = 0$: From (201) $x_5 = 0$ because $x_4 \neq 0$ at steady state. Consequently, $x_6 = \overline{T} \neq 0$ and from (202) it follows that $x_7 = 0$. Thus $x_8 = \overline{R} \neq 0$ at steady state.

- $k_{h2} = k_{4r} = 0$: From (202) either $x_6 = 0$ or $x_7 = 0$. If $x_6 = 0$ we have that $x_5 \neq 0$. From (201) if follows that $x_4 = 0$, which is a contradiction. Therefore $x_7 = 0$, and hence $x_8 = \overline{R} \neq 0$ at steady state.

Assume now that none of the two scenarios above occur, and further that $k_{h1} = k_{h2} = 0$ does not occur. If $x_9 = 0$ is a solution at steady state, then by (203)

we must either have $x_1 = 0$ or $x_4 = 0$. If $x_1 = 0$ then $x_2 \neq 0$. Further from (199) we have $k_{h1}x_4 = k_{h2}x_8 = 0$. Hence from (200) we have $x_3 = 0$ and as a consequence $x_4 \neq 0$. If $k_{h1} \neq 0$ then $k_{h1}x_4 \neq 0$, which is a contradiction. Hence, assume that $k_{h1} = 0$. Then using (201) we deduce that $x_5 = 0$ and hence $x_6 \neq 0$ from the conservation law. From (202) we obtain $x_7 = 0$ which contradicts (198) for $\overline{R} > 0$ only if $k_{h2} \neq 0$.

Assume now that $x_4 = 0$. Then by (196) we have $x_3 \neq 0$. Further, from (201) we have $k_{h2}x_8 = 0$. From (200) we have $x_2 = 0$ and from (195) $x_1 \neq 0$, contradicting (199).

Therefore, if $k_{h1} \neq 0$ or $k_{h2} \neq 0$ then $x_9 = 0$ is not a solution at steady state. If one of the concentrations $x_1, \ldots, x_4$ is zero at steady state, then the positive term in one of the equations (200),(201),(203) vanishes, implying that all the other monomials must vanish as well. For any of the equations, it would follow imply that $k_6x_9 = 0$ contradicting $x_9 \neq 0$. That is, $x_1, \ldots, x_4 \neq 0$ at steady state. If $x_5 = 0$ and $k_{3r} \neq 0$ then using (201) and (197), $x_3 = 0$ which is a contradiction. If $k_{3r} = 0$ then $k_{h2} \neq 0$ (by assumption) and hence $x_8 = 0$. By (202) we have $x_6 = 0$ or $x_7 = 0$. The latter contradicts (198) because $x_8 = 0$. Hence $x_6 = 0$. However this contradicts (197), because $x_5 = 0$.

Therefore, if $k_{h2} \neq 0$ or if $k_{h2} = 0$ but $k_{3r}k_{4r}k_{h1} \neq 0$, then there are no zero concentrations at steady state.

C.2.2 Steady-state relations

Here we derive the expressions shown in (205). We study the concentrations at steady state that are non-zero. We assume either (1) $k_{h2} \neq 0$ or (2) $k_{h2} = 0$ and $k_{3r}k_{4r}k_{h1} \neq 0$.

(1) Using the total amount equation for $\overline{R}$, we have

$$x_7 = \overline{R} - x_8, \tag{208}$$

such that x_7 is expressed as a decreasing function of x_8. We have $x_7, x_8 > 0$ if and only if $0 < x_8 < \alpha_1 := \overline{R}$.

(2) Using (202) and the total amount equation for $\overline{T}$ we obtain

$$x_6 = \frac{x_8(k_{4r}\overline{T} + k_{h2})}{k_4x_7 + k_{4r}x_8}, \qquad x_5 = \frac{k_4\overline{T}x_7 - k_{h2}x_8}{k_4x_7 + k_{4r}x_8}. \tag{209}$$

The expression for x_5 decreases in x_8 and increases in x_7. Since x_7 decreases in x_8, we conclude that after substituting x_7 with (208), x_5 decreases in x_8. Similarly x_6 increases in x_8.

For $x_5, x_6, x_7 > 0$, we require $k_4\overline{T}(\overline{R} - x_8) - k_{h2}x_8 > 0$, that is,

$$0 < x_8 < \alpha_2 := \frac{k_4\overline{T} \cdot \overline{R}}{k_4\overline{T} + k_{h2}} \leq \alpha_1.$$

Hence, $0 < x_8 < \alpha_2$ if and only if $x_5, x_6, x_7, x_8 > 0$.

(3) Using (201) and the total amount equation for $\overline{C}$, we obtain:

$$x_4 = \frac{k_{h2}x_8 + k_{3r}x_6(C - x_9)}{k_3x_5 + k_{3r}x_6}, \qquad x_3 = \frac{k_3x_5(C - x_9) - k_{h2}x_8}{k_3x_5 + k_{3r}x_6}. \tag{210}$$

x_4 is positive provided $0 < x_8 < \alpha_2$ and $0 < x_9 < C$. x_3 is positive provided x_8, x_9 satisfy $k_3x_5(C - x_9) > k_{h2}x_8$, that is,

$$x_9 < \frac{k_3x_5C - k_{h2}x_8}{k_3x_5} \leq C.$$

The right-hand side decreases in x_8. It is zero when $k_3x_5C = k_{h2}x_8$. If $x_8 > 0$ then $x_5 > 0$ and hence the value that makes the right-hand side zero satisfies $x_8 < \alpha_2$. Therefore, $x_3, \ldots, x_9$ are positive for x_8, x_9 in the set

$$\Omega_1 := \left\{(x_8, x_9) \in \mathbb{R}^2_+ | x_9 < \frac{k_3x_5C - k_{h2}x_8}{k_3x_5}\right\},$$

with x_5 given by (208) and (209). For each value of $0 < x_9 < C$, let $\beta(x_9)$ be the value of x_8 for which $x_9 = \frac{k_3x_5C - k_{h2}x_8}{k_3x_5}$, that is, the upper-bound of allowed values for x_8. Note that $\beta(x_9)$ decreases in x_9.

The expression for x_4 in (210) increases in x_8 and decreases in x_5, x_9. The derivative of x_4 with respect to x_6 equals

$$\frac{k_{3r}x_3}{(k_3x_5 + k_{3r}x_6)},$$

and hence it is positive provided $x_3, x_5, x_6 > 0$. Therefore, for $(x_8, x_9) \in \Omega_1$, x_4 increases in x_8 and decreases in x_9. Similarly, x_3 decreases in x_8 and in x_9.

(4) Using (200) and the total amount equation for $\overline{H}$ we obtain

$$x_1 = \frac{k_2x_3(\overline{H} - x_9) - k_6x_9 - k_{h1}x_4 - k_{h2}x_8}{k_2x_3 + k_{2r}x_4}, \quad x_2 = \frac{k_{2r}x_4(\overline{H} - x_9) + k_6x_9 + k_{h1}x_4 + k_{h2}x_8}{k_2x_3 + k_{2r}x_4}.$$

For $(x_8, x_9) \in \Omega_1$, x_1 is positive provided that

$$k_2 x_3(\overline{H} - x_9) > k_{h1} x_4 + k_{h2} x_8 + k_6 x_9.$$

Fix a value of $0 < x_9 < \min(\overline{H}, \overline{C})$. Then the left-hand side of the inequality is a decreasing function of x_8 and the right-hand side of the inequality is increasing in x_8. It follows that there exists a value $\gamma(x_9)$ such that the inequality is fulfilled if and only if $x_8 < \gamma(x_9)$.

If $x_8 = \beta(x_9)$, then $x_3 = 0$ while the right-hand side of the inequality is positive. It follows that $\beta(x_9) > \gamma(x_9)$. x_2 is positive if x_4, x_3 are positive and $x_9 < H$. Therefore, $x_1, \ldots, x_9$ are positive provided x_8, x_9 belong to

$$\Omega_2 := \{(x_8, x_9) \in \mathbb{R}^2_+ |\ x_9 < \min(\overline{H}, \overline{C}),\ k_2 x_3(\overline{H} - x_9) > k_{h1} x_4 + k_{h2} x_8 + k_6 x_9\}.$$

It can be seen that in Ω_2, x_1 decreases in x_9 and in x_8. Similarly, x_2 increases in x_8. Further, the numerator of x_1 also decreases in x_9. It follows that the supremum of x_8 in Ω_2 is obtained by setting $x_9 = 0$:

$$k_2 x_3 \overline{H} - k_{h1} x_4 - k_{h2} x_8 = 0,$$

where $x_9 = 0$ is inserted into the expression of x_3, x_4. The solution of this equation is precisely the value α obtained in the monofunctional case (see equation 57). Furthermore, the possible values of x_8 in Ω_2 are in $I = (0, \alpha)$.

(5) Using (203) we obtain another expression for x_1 at steady state:

$$x_1 = \frac{(k_{5r} + k_6) x_9}{k_5 x_4} = \frac{x_9}{k_y x_4},$$

where $k_y = k_5/(k_{5r} + k_6)$. This expression decreases in x_8. We equate the two expressions for x_1:

$$\frac{x_9}{k_y x_4} = \frac{k_2 x_3 \overline{H} - k_2 x_3 x_9 - k_6 x_9 - k_{h1} x_4 - k_{h2} x_8}{k_2 x_3 + k_{2r} x_4} \tag{211}$$

in order to relate x_8 and x_9. This equality does not provide a linear equation in x_8 nor in x_9, when substituting the expressions for x_3, x_4 in terms of x_8, x_9. Thus, we have to proceed in a different way from what we have done so far. For a fixed value of x_8 in I, the left-hand side of the equation increases in x_9 and the right-hand side decreases in x_9. Therefore, for a fixed value of x_8 in $I = (0, \alpha)$, the two sides of the equality intersect in exactly one point: $x_9 = g(x_8)$. Since the

intersection point ensures that the right-hand side is positive, the intersection point satisfies by construction that $(x_8, g(x_8)) \in \Omega_2$.

We do not have an analytical description of g but we have a procedure to determine $g(x_8)$ from a given x_8. The function g is given by the Implicit Function Theorem. Let

$$G(x_8, x_9) = (k_2 x_3(\overline{H} - x_9) - k_6 x_9 - k_{h1} x_4 - k_{h2} x_8) k_y x_4 - x_9(k_2 x_3 + k_{2r} x_4) = 0.$$

Then, for every value of x_8, $g(x_8)$ is the first positive root of $G(x_8, x_9)$. It follows that g is continuous in I and differentiable. The derivative of g with respect to x_8 is given by

$$g'(x_8) = -\frac{(\partial G/\partial x_8)(x_8, g(x_8))}{(\partial G/\partial x_9)(x_8, g(x_8))}.$$

The function g can be extended to $x_8 = 0$ with $g(0) = 0$.

C.2.3 Signal-response curve

The entries of (205) are derived using all steady-state equations except for (199). From (199) we obtain that

$$k_s = f_b(x_8) = \frac{k_{h1} x_4 + k_{h2} x_8 + k_6 x_9}{x_1}. \tag{212}$$

If $x_8 \in I$, then $f_b(x_8)$ is positive. Therefore, all concentrations at steady state are positive. This function is continuous and differentiable. When x_8 approaches the upper bound of the interval I, α, then x_1 tends to zero, x_4 to some finite number and x_9 to zero. Hence k_s grows to infinity (provided k_{h1} or k_{h2} are non-zero, see below for the case $k_{h1} = k_{h2} = 0$). It follows that the image of f_b is $(0, +\infty)$ which guarantees the existence of at least one steady state. The function f_b can be differentiably extended at zero such that $f_b(0) = 0$. Using the Chemical Reaction Network Toolbox (`https://crnt.osu.edu/LecturesOnReactionNetworks`), we know that the system does not admit multiple positive steady states. By continuity, it follows that f_b must be monotone, that is, an increasing function. By the Inverse Function Theorem, there exists a continuous and differentiable function in $(0, +\infty)$,

$$x_8 = \varphi_b(k_s)$$

defined by $\varphi_b(k_s) = x_8$ if and only if $k_s = f_b(x_8)$.

If we increase k_y while keeping k_6 fixed, the right-hand side of (211) increases. It

follows that the value $g(x_8)$ increases. Consequently, x_4 increases and x_1 decreases, which implies that k_s must increase as well. We conclude that as k_y increases the graphs of f_b pile on top of each other and hence the graphs of φ_b lie below each other.

$k_{h1} = k_{h2} = 0$**:** First of all, an easy check shows that $(0, \overline{H}, 0, \overline{C}, 0, \overline{T}, 0, \overline{R}, 0)$ is a steady state for all values of k_s. But for k_s small enough, a second positive steady state exists as well.
In this case we have $\alpha = \overline{R}$ and

$$k_s = f_b(x_8) = \frac{k_6 x_9}{x_1}.$$

Hence both the numerator and denominator of f_b tend to zero as x_8 tend to $\alpha = \overline{R}$. By plugging the expression of x_1 into f_b, we have:

$$f_b(x_8) = \frac{k_6 k_{3r} k_{4r} k_y (\overline{C} - x_9) x_8}{k_3 k_4 (\overline{R} - x_8) - k_{3r} k_{4r} x_8}.$$

We deduce that when $x_8 = \overline{R}$, then $x_9 = 0$ and $f_b(x_8) = k_6 k_y \overline{C}$. It follows that $k_s = f_b(x_8)$ does not tend to infinity as x_8 approaches the upper bound of I, $\overline{R}$. In this case, the signal-response curve is defined by f_b^{-1} for $k_s \in [0, k_6 k_y \overline{C}]$ and is constant at $\overline{R}$ for $k_s > k_6 k_y \overline{C}$.

D

APPENDIX

D.1 PRISM script for the generic phosphorelay model discussed in section 2.2.2.2

```
ctmc

const int HKInit;
const int HKpInit = 0;
const int RECInit;
const int RECpInit = 0;
const int HptInit;
const int HptpInit = 0;
const int RRInit;
const int RRpInit = 0;

const int HKTot = HKInit+HKpInit;
const int RECTot = RECInit+RECpInit;
const int HptTot = HptInit+HptpInit;
const int RRTot = RRInit+RRpInit;
```

```
const double ks;
const double k2=9.99e-1;
const double k3=2.91e-1;
const double k4=4.14e-1;
const double k5=0.0178;
const double k2r=0;
const double k3r=1.16;
const double k4r=1.41;
const double k2h=6.64e-2;

module PR4r

HK : [0..HKTot] init HKInit;
HKp : [0..HKTot] init HKpInit;
REC : [0..RECTot] init RECInit;
RECp : [0..RECTot] init RECpInit;
Hpt : [0..HptTot] init HptInit;
Hptp : [0..HptTot] init HptpInit;
RR : [0..RRTot] init RRInit;
RRp : [0..RRTot] init RRpInit;

// HK -> HKp
[ks] ks>0 & HK>0 -> (ks*HK): \
(HK'=HK-1) & (HKp'=min(HKp+1, HKTot));

// HKp + REC -> HK + RECp
[k2] k2>0 & HKp>0 & REC>0 -> (k2*HKp*REC):
(HKp'=HKp-1) & (REC'=REC-1) & (HK'=min(HK+1, HKTot))&
(RECp'=min(RECp+1, RECTot));

// RECp + Hpt -> REC + Hptp
[k3] k3>0 & RECp>0 & Hpt>0 -> (k3*RECp*Hpt):
(RECp'=RECp-1) & (Hpt'=Hpt-1) & (REC'=min(REC+1, RECTot)) &
(Hptp'=min(Hptp+1, HptTot));

// Hptp + RR -> Hpt + RRp
[k4] k4>0 & Hptp>0 & RR>0 -> (k4*Hptp*RR) :
(Hptp'=Hptp-1) & (RR'=RR-1) & (Hpt'=min(Hpt+1, HptTot)) &
(RRp'=min(RRp+1, RRTot));
```

```
// RRp -> RR
[k5] k5>0 & RRp>0 -> (k5*RRp) : (RRp'=RRp-1) &
(RR'=min(RR+1, RRTot));

// HK + RECp -> HKp + REC
[k2r] k2r>0 & HK>0 & RECp>0 -> (k2r*HK*RECp) :
(HK'=HK-1) & (RECp'=RECp-1)& (HKp'=min(HKp+1, HKTot)) &
(REC'=min(REC+1, RECTot));

// REC + Hptp -> RECp + Hpt
[k3r] k3r>0 & REC>0 & Hptp>0 -> (k3r*REC*Hptp) :
(REC'=REC-1) & (Hptp'=Hptp-1)& (RECp'=min(RECp+1, RECTot)) &
(Hpt'=min(Hpt+1, HptTot));

// Hpt + RRp -> Hptp + RR
[k4r] k4r>0 & Hpt>0 & RRp>0 -> (k4r*Hpt*RRp) : (Hpt'=Hpt-1) &
(RRp'=RRp-1)& (Hptp'=min(Hptp+1, HptTot)) &
(RR'=min(RR+1, RRTot));

// RECp -> REC
[k2h] k2h>0 & RECp>0 -> (k2h*RECp) :
(RECp'=RECp-1) & (REC'=min(REC+1, RECTot));

endmodule

rewards "HK" true : HK; endrewards
rewards "HKp" true : HKp; endrewards
rewards "REC" true : REC; endrewards
rewards "RECp" true : RECp; endrewards
rewards "Hpt" true : Hpt; endrewards
rewards "Hptp" true : Hptp; endrewards
rewards "RR" true : RR; endrewards
rewards "RRp" true : RRp; endrewards

rewards "HK_sq" true : HK*HK; endrewards
rewards "HKp_sq" true : HKp*HKp; endrewards
rewards "REC_sq" true : REC*REC; endrewards
rewards "RECp_sq" true : RECp*RECp; endrewards
```

```
rewards "Hpt_sq" true : Hpt*Hpt; endrewards
rewards "Hptp_sq" true : Hptp*Hptp; endrewards
rewards "RR_sq" true : RR*RR; endrewards
rewards "RRp_sq" true : RRp*RRp; endrewards

rewards "time" true : 1; endrewards
```

D.2 Converting kinetic rates to probabilistic parameters in PRISM

The table below shows how kinetic rate parameters and protein concentrations from the ODE model discussed in section 2.2.1 can be converted to probabilistic rates and molecule numbers for use in PRISM.

Parameters	Mass Action Values	Dimension	Scaling by g= 53191(M^{-1})	Stochastic Values	Dimension
k_s	0 .. 0.1	s^{-1}	/1	0 .. 0.1	s^{-1}
k_2	3.96E+04	$M^{-1}s^{-1}$	/g	7.45E-01	s^{-1}
k_3	6.59E+04	$M^{-1}s^{-1}$	/g	1.24E+00	s^{-1}
k_4	1.63E+03	$M^{-1}s^{-1}$	/g	3.06E-02	s^{-1}
k_5	0.064663803	s^{-1}	/1	0.017816822	s^{-1}
k_{2r}	0	$M^{-1}s^{-1}$	/g	0.064663803	s^{-1}
k_{3r}	8.84E+04	$M^{-1}s^{-1}$	/g	1.66E+00	s^{-1}
k_{4r}	2.69E+04	$M^{-1}s^{-1}$	/g	5.05E-01	s^{-1}
k_{2h}	0.092685844	s^{-1}	/1	9.27E-02	s^{-1}
HK_{tot}	1.76E-04	M	*g	9.38E+00	1
REC_{tot}	1.76E-04	M	*g	9.38E+00	1
Hpt_{tot}	1.76E-04	M	*g	9.38E+00	1
RR_{tot}	1.76E-04	M	*g	9.38E+00	1

Table D.1: Shows the conversion of kinetic rates used in ODE models for topology 30 (in the sigmoidal regime) to probabilistic rates used in the PRISM model.

Parameters	Mass Action Values	Dimension	Scaling by g= 53191(M^{-1})	Stochastic Values	Dimension
k_s	0 .. 0.1	s^{-1}	/1	0 .. 0.1	s^{-1}
k_2	6.16E+04	$M^{-1}s^{-1}$	/g	1.16E+00	s^{-1}
k_3	3.26E+03	$M^{-1}s^{-1}$	/g	6.12E-02	s^{-1}
k_4	1.72E+04	$M^{-1}s^{-1}$	/g	3.24E-01	s^{-1}
k_5	0	s^{-1}	/1	0	s^{-1}
k_{2r}	0	$M^{-1}s^{-1}$	/g	0	s^{-1}
k_{3r}	2.31E+04	$M^{-1}s^{-1}$	/g	4.35E-01	s^{-1}
k_{4r}	2.55E+04	$M^{-1}s^{-1}$	/g	4.79E-01	s^{-1}
k_{2h}	4.15E-02	s^{-1}	/1	4.15E-02	s^{-1}
HK_{tot}	1.84E-04	M	*g	9.79E+00	1
REC_{tot}	1.84E-04	M	*g	9.79E+00	1
Hpt_{tot}	1.84E-04	M	*g	9.79E+00	1
RR_{tot}	1.84E-04	M	*g	9.79E+00	1

Table D.2: Shows the conversion of kinetic rates used in ODE models for topology 14 (in the sigmoidal regime) to probabilistic rates used in the PRISM model.

Parameters	Mass Action Values	Dimension	Scaling by g= 53191(M^{-1})	Stochastic Values	Dimension
k_s	0 .. 0.1	s^{-1}	/1	0 .. 0.1	s^{-1}
k_2	6.99E+04	$M^{-1}s^{-1}$	/g	1.31E+00	s^{-1}
k_3	8.96E+04	$M^{-1}s^{-1}$	/g	1.68E+00	s^{-1}
k_4	4.88E+04	$M^{-1}s^{-1}$	/g	9.18E-01	s^{-1}
k_5	0	s^{-1}	/1	0	s^{-1}
k_{2r}	0	$M^{-1}s^{-1}$	/g	0	s^{-1}
k_{3r}	3.71E+04	$M^{-1}s^{-1}$	/g	6.98E-01	s^{-1}
k_{4r}	3.91E+04	$M^{-1}s^{-1}$	/g	7.34E-01	s^{-1}
k_{2h}	9.23E-02	s^{-1}	/1	9.23E-02	s^{-1}
HK_{tot}	1.04E-04	M	*g	5.55E+00	1
REC_{tot}	1.04E-04	M	*g	5.55E+00	1
Hpt_{tot}	1.04E-04	M	*g	5.55E+00	1
RR_{tot}	1.04E-04	M	*g	5.55E+00	1

Table D.3: Shows the conversion of kinetic rates used in ODE models for topology 14 (in the hyperbolic regime) to probabilistic rates used in the PRISM model.

E

APPENDIX

E.1 Proofs showing the existence of $2n+1$ positive steady states as discussed in section 3.2.4.3

E.1.1 Proof of Lemma 1

To simplify the notation, we prove that for all $\alpha_1, \ldots, \alpha_5 > 0$ there exist $k_1, \ldots, k_5, H > 0$ such that

$$\alpha_1 = k_1 k_2 k_4 k_5 H \qquad \alpha_2 = k_1 k_2 k_3 k_5 H \tag{213}$$

$$\alpha_3 = k_1 k_4 k_5 + k_2 k_4 k_5 \qquad \alpha_4 = k_1 k_2 k_5 + k_1 k_3 k_5 \qquad \alpha_5 = k_1 k_2 k_3. \tag{214}$$

Using (214) we solve iteratively for k_3, k_5 and k_4 and obtain:

$$k_3 = \frac{\alpha_5}{k_1 k_2}, \qquad k_5 = \frac{\alpha_4 k_2}{k_1 k_2^2 + \alpha_5}, \qquad k_4 = \frac{\alpha_3 (k_1 k_2^2 + \alpha_5)}{\alpha_4 k_2 (k_1 + k_2)}. \tag{215}$$

Using the first equation in (213), we have

$$H = \frac{\alpha_2}{k_1 k_2 k_4 k_5} = \frac{\alpha_2 (k_1 + k_2)}{\alpha_3 k_1 k_2}. \tag{216}$$

Finally, using the second equation in (213), we have

$$0 = \alpha_2 - k_1 k_2 k_3 k_5 H = \alpha_2 - \frac{\alpha_2(k_1+k_2)\alpha_4\alpha_5}{\alpha_3 k_1(k_1 k_2^2 + \alpha_5)},$$

which is equivalent to

$$0 = \alpha_2\alpha_3 k_2^2 k_1^2 + \alpha_2\alpha_5(\alpha_3 - \alpha_4)k_1 - \alpha_2\alpha_4\alpha_5 k_2. \tag{217}$$

Fix any $k_2 > 0$. Since the polynomial (217) is a polynomial in k_1, with positive leading coefficient and negative independent term, it has a unique positive real root. As a consequence, for any $k_2 > 0$, (213) and (214) hold with k_1 defined such that (217) holds, and $k_3, k_4, k_5, H > 0$ as in (215), (216). □

E.1.2 Proof of Lemma 2

We have that

$$\begin{aligned} p(x) &= (x-1)\prod_{i=1}^{n}(x^2+b_i) + \sum_{i=1}^{n}\Big(a_i x \prod_{j\neq i}(x^2+b_j)\Big) \\ &= \prod_{i=1}^{n}(x^2+b_i) + x\Big(\prod_{i=1}^{n}(x^2+b_i) + \sum_{i=1}^{n}\big(a_i\prod_{j\neq i}(x^2+b_j)\big)\Big). \end{aligned}$$

If we write

$$p_1(x) = \prod_{i=1}^{n}(x^2+b_i) = \beta_0 x^{2n} + \cdots + \beta_{2n-2}x^2 + \beta_{2n},$$

then a standard computation shows that

$$\beta_{2k} = \sum_{\{j_1,\ldots,j_k\}\subseteq[n]}\prod_{\ell=1}^{k} b_{j_\ell}, \qquad \text{and} \qquad \beta_{2k+1} = 0,$$

for $k = 0, 1, \ldots, n$. In particular, only coefficients of even degree are nonzero. The second summand of $p(x)$,

$$p_2(x) = x\Big(\prod_{i=1}^{n}(x^2+b_i) + \sum_{i=1}^{n}\big(a_i\prod_{j\neq i}(x^2+b_j)\big)\Big),$$

has only terms of odd degree. Therefore, $c_{2k+1} = \alpha_{2k}$ as claimed and c_{2k} is the coefficient of degree $2n+1-2k$ of $p_2(x)$. Note that the polynomial $p_2(x)$ can be

written as:

$$p_2(x) = xp_1(x) + x\sum_{i=1}^{n}\big(a_i\prod_{j\neq i}(x^2+b_j)\big).$$

As above, if we write $\prod_{j\neq i}^{n}(x^2+b_j) = \gamma_{i,0}x^{2n-2} + \cdots + \gamma_{i,2n-4}x^2 + \gamma_{i,2n-2}$, then

$$\gamma_{i,2k} = \sum_{\{j_1,\ldots,j_k\}\subseteq[n]\setminus\{i\}}\prod_{\ell=1}^{k} b_{j_\ell}, \qquad \text{and} \qquad \gamma_{2k+1} = 0,$$

for $k = 0, 1, \ldots, n-1$. The coefficient of degree $2n+1-2k$ of $p(x)$ is then

$$c_{2k} = \beta_{2k} + \sum_{i=1}^{n} a_i\gamma_{i,2k-2},$$

as claimed. □

E.1.3 Proof of Lemma 3

To simplify the presentation, we denote by $\mathcal{P}_k(n)$ the set of all subsets of $[n]$ with k elements. Similarly, given $i \in [n]$, we denote by $\mathcal{P}_{k-1}(n,i)$ the set of all subsets of $[n]$ with $k-1$ elements that do not contain i. We want to show that if we define $b_i = \frac{a_i^2}{4}$ for $i = 1, \ldots, n$, then,

$$c_{2k}^2 - 4c_{2k-1}c_{2k+1} > 0$$

for all $k = 1, \ldots, n$. Recall that we defined

$$c_{2k+1} = -\sum_{L\in\mathcal{P}_k(n)}\prod_{\ell\in L} b_\ell, \qquad c_{2k} = \sum_{L\in\mathcal{P}_k(n)}\prod_{\ell\in L} b_\ell + \sum_{i=1}^{n} a_i \sum_{J\in\mathcal{P}_{k-1}(n,i)}\prod_{j\in J} b_j.$$

It is clear that if $k \geq 1$, then

$$\begin{aligned} c_{2k}^2 &> \Big(\sum_{i=1}^{n} a_i \sum_{J\in\mathcal{P}_{k-1}(n,i)}\prod_{j\in J} b_j\Big)^2 \\ &> \sum_{i=1}^{n}\sum_{J\in\mathcal{P}_{k-1}(n,i)} a_i^2\prod_{j\in J} b_j^2 + 2\sum_{i=1}^{n}\sum_{\substack{J,L\in\mathcal{P}_{k-1}(n,i)\\ J\neq L}} a_i^2\prod_{j\in J} b_j\prod_{\ell\in L} b_\ell. \end{aligned} \tag{218}$$

On the other hand,

$$4c_{2k+1}c_{2k-1} = 4\left(\sum_{L\in\mathcal{P}_k(n)}\prod_{\ell\in L} b_\ell\right)\left(\sum_{J\in\mathcal{P}_{k-1}(n)}\prod_{j\in J} b_j\right) = 4\sum_{\substack{L\in\mathcal{P}_k(n)\\ J\in\mathcal{P}_{k-1}(n)}}\left(\prod_{\ell\in L} b_\ell\prod_{j\in J} b_j\right)$$
$$= 4\left(\sum_{i=1}^{n}\sum_{J\in\mathcal{P}_{k-1}(n,i)} b_i\prod_{j\in J} b_j^2\right) + 4\sum_{\substack{L\in\mathcal{P}_k(n)\\ J\in\mathcal{P}_{k-1}(n),J\not\subseteq L}}\left(\prod_{\ell\in L} b_\ell\prod_{j\in J} b_j\right). \quad (219)$$

If we set $b_i = \frac{a_i^2}{4}$, then the first two summands of (218) and (219) agree:

$$\sum_{i=1}^{n}\sum_{J\in\mathcal{P}_{k-1}(n,i)} a_i^2\prod_{j\in J} b_j^2 = 4\sum_{i=1}^{n}\sum_{J\in\mathcal{P}_{k-1}(n,i)}\frac{a_i^2}{4}\prod_{j\in J} b_j^2 = 4\left(\sum_{i=1}^{n}\sum_{J\in\mathcal{P}_{k-1}(n,i)} b_i\prod_{j\in J} b_j^2\right).$$

We let

$$B_1 = \sum_{i=1}^{n}\sum_{\substack{J,L\in\mathcal{P}_{k-1}(n,i)\\ J\neq L}} a_i^2\prod_{j\in J} b_j\prod_{\ell\in L} b_\ell = 4\sum_{i=1}^{n}\sum_{\substack{J,L\in\mathcal{P}_{k-1}(n,i)\\ J\neq L}} b_i\prod_{j\in J} b_j\prod_{\ell\in L} b_\ell$$
$$B_2 = 2\sum_{\substack{L\in\mathcal{P}_k(n)\\ J\in\mathcal{P}_{k-1}(n),J\not\subseteq L}}\left(\prod_{\ell\in L} b_\ell\prod_{j\in J} b_j\right).$$

Then, to show that $c_{2k}^2 > 4c_{2k+1}c_{2k-1}$ it is enough to show that $\frac{B_1}{4} \geq \frac{B_2}{2}$. To this end, observe that given any pair of sets $L \in \mathcal{P}_k(n)$ and $J \in \mathcal{P}_{k-1}(n)$, such that $J \not\subseteq L$, we can choose $i \in L$ such that $L' := L \setminus \{i\}$ satisfies $L' \neq J$. Therefore, every summand of B_2 is a summand of B_1 as well. This finishes the proof. □

E.1.4 Proof of Lemma 4

Let $\delta_k = c_{2k+1}^2 - 4c_{2k}c_{2k+2}$. For $M > 0$ and $a_1 > 0$, we let $a_i = \frac{a_n}{M^{i-1}}$ and $b_i = \frac{a_i^2}{4}$. Therefore, we have that

$$b_i = \frac{a_1^2}{4M^{2(i-1)}},$$

for $i = 1, \ldots, n$. With these substitutions, $\delta_k = \delta_k(a_1)$ becomes a polynomial in a_1. If there exists a choice of $M > 0$ such that the coefficient of *smallest degree* of δ_k is positive for all k, then $\delta_k(a_1) > 0$, for a_1 small enough and all k. Therefore, we compute the coefficient of smallest degree of δ_k in a_1.

We use the notation introduced in the proof of Lemma 3. By definition, we have that

$$c_{2k+1}^2(a_1) = \left(\sum_{J \in \mathcal{P}_k(n)} \prod_{j \in J} \frac{a_1^2}{4M^{2(j-1)}} \right)^2 = a_1^{4k} \left(\sum_{J \in \mathcal{P}_k(n)} \prod_{j \in J} \frac{1}{4M^{2(j-1)}} \right)^2$$

$$\begin{aligned} c_{2k}(a_1) &= \sum_{J \in \mathcal{P}_k(n)} \prod_{j \in J} \frac{a_1^2}{4M^{2(j-1)}} + \sum_{i=1}^{n} \frac{a_1}{M^{i-1}} \sum_{J \in \mathcal{P}_{k-1}(n,i)} \prod_{j \in J} \frac{a_1^2}{4M^{2(j-1)}} \\ &= a_1^{2k} \left(\sum_{J \in \mathcal{P}_k(n)} \prod_{j \in J} \frac{1}{4M^{2(j-1)}} \right) + a_1^{2k-1} \left(\sum_{i=1}^{n} \frac{1}{M^{i-1}} \sum_{J \in \mathcal{P}_{k-1}(n,i)} \prod_{j \in J} \frac{1}{4M^{2(j-1)}} \right) \end{aligned}$$

The polynomial $c_{2k+1}^2(a_1)$ consists of one term of degree $4k$ in a_1. The polynomial $c_{2k}(a_1)$ is a sum of a term of degree $2k$ and one of degree $2k-1$. Similarly, the polynomial $c_{2k+2}(a_1)$ is a sum of a term of degree $2k+2$ and one of degree $2k+1$. Hence, the product $c_{2k}c_{2k+2}$ is a polynomial with lowest degree $4k$. If the coefficient of degree $4k$ of δ_k is nonzero, then it is the coefficient if smallest degree. By denoting the coefficient of degree $4k$ by β_k, we have:

$$\begin{aligned} \beta_k = &\left(\sum_{J \in \mathcal{P}_k(n)} \prod_{j \in J} \frac{1}{4M^{2(j-1)}} \right)^2 \\ &- 4 \left(\sum_{i=1}^{n} \frac{1}{M^{i-1}} \sum_{J \in \mathcal{P}_{k-1}(n,i)} \prod_{j \in J} \frac{1}{4M^{2(j-1)}} \right) \left(\sum_{s=1}^{n} \frac{1}{M^{s-1}} \sum_{L \in \mathcal{P}_{k+1}(n,s)} \prod_{\ell \in L} \frac{1}{4M^{2(\ell-1)}} \right). \end{aligned}$$

For $M > 1$, the largest summand in the positive summand of β_k is given by the choice of set $J = \{1, \ldots, k\} \in \mathcal{P}_k(n)$ and takes the value

$$X_1 := \prod_{j=1}^{k} \frac{1}{16M^{4(j-1)}} = \frac{1}{4^{2k} M^{4\sum_{\ell=0}^{k-1} \ell}} = \frac{1}{4^{2k} M^{2k^2-2k}}.$$

The largest summand in the negative summand of β_k is given by the choice $i = k, s = k+2$, $J = \{1, \ldots, k-1\} \in \mathcal{P}_{k-1}(n,i)$, and $L = \{1, \ldots, k+1\} \in \mathcal{P}_{k+1}(n,s)$. Proceeding as above, the term takes the value

$$X_2 := 4 \left(\frac{1}{M^{k-1}} \frac{1}{4^{k-1} M^{(k-1)(k-2)}} \right) \left(\frac{1}{M^{k+1}} \frac{1}{4^{k+1} M^{(k+1)k}} \right) = \frac{1}{4^{2k-1} M^{2k^2+2}}.$$

For M large enough, $X_1 > X_2$. It follows that for M large, X_1 dominates, and hence β_k is positive. This finishes the proof. □

F

APPENDIX

F.1 Parameters used to generate all figures in Chapter 3

Constant	Fig 3.2		3.6B			Fig 3.6D	Fig 3.8[b]			Fig 3.5	Fig 3.7	Reactions
	b	c[a]	$n=1$	$n=2$	$n=3$		B	C	D			
$k_{1,1}$	0 .. 1.5	0.7329	1	0.1	0.591	0 .. 1.4	0 .. 2	0 .. 2	0 .. 2	0 .. 2.5	0 .. 5	$HK^1_{00} \rightarrow HK^1_{P0}$
$k_{1,2}$	100	100	10	120	4050	10	100	100	100	100	100	$HK^1_{P0} \rightarrow HK^1_{0P}$
$k_{1,3}$	$100 \times k_{1,1}$	73.29	15	17.95	127.2	15	$80 \times k_{1,1}$	$100 \times k_{1,1}$	$100 \times k_{1,1}$	$100 \times k_{1,1}$	$10 \times k_{1,1}$	$HK^1_{0P} \rightarrow HK^1_{PP}$
$k_{1,4}$	50	50	30	0.1795	5.986	0.1	50	50	50	50	50	$HK^1_{0P} + Hpt^1 \rightarrow HK^1_{00} + Hpt^1_P$
$k_{1,5}$	100	100	200	0.713	16.7	50	100	100	100	100	100	$HK^1_{PP} + Hpt^1 \rightarrow HK^1_{P0} + Hpt^1_P$
$k_{1,6}$	5	5	20	1	100	-	-	-	-	-	-	$Hpt^1_P \rightarrow Hpt^1$
$k_{1,7}$	-	-	-	-	-	-	10	10	10	10	-	$HK^1_{00} + Hpt^1_P \rightarrow HK^1_{0P} + Hpt^1$
$k_{1,8}$	-	-	-	-	-	-	10	10	10	10	-	$HK^1_{P0} + Hpt^1_P \rightarrow HK^1_{PP} + Hpt^1$
$k_{1,9}$	-	-	-	-	-	-	10	10	10	10	-	$HK^1_{0P} \rightarrow HK^1_{00}$
$k_{1,10}$	-	-	-	-	-	-	10	10	10	10	-	$HK^1_{PP} \rightarrow HK^1_{P0}$
$k_{1,11}$	-	-	-	-	-	0.1	100	100	100	100	1	$Hpt^1_P + RR \rightarrow Hpt^1 + RR_P$
$k_{1,12}$	-	-	-	-	-	-	100	100	100	100	-	$Hpt^1 + RR_P \rightarrow Hpt^1_P + RR$
$k_{2,1}$	-	-	-	0.002	0.363	$k_{1,1}$	0 .. 2	0 .. 2	0 .. 2	-	$k_{1,1}$	$HK^2_{00} \rightarrow HK^2_{P0}$
$k_{2,2}$	-	-	-	500	6500	200	100	100	100	-	100	$HK^2_{P0} \rightarrow HK^2_{0P}$
$k_{2,3}$	-	-	-	160	24	50	$80 \times k_{2,1}$	$100 \times k_{2,1}$	$100 \times k_{2,1}$	-	$10 \times k_{1,1}$	$HK^2_{0P} \rightarrow HK^2_{PP}$
$k_{2,4}$	-	-	-	0.147	3.633	-	50	50	50	-	-	$HK^2_{0P} + Hpt^1 \rightarrow HK^2_{00} + Hpt^1_P$
	-	-	-	-	-	1	-	-	-	-	50	$HK^2_{0P} + Hpt^2 \rightarrow HK^2_{00} + Hpt^2_P$
$k_{2,5}$	-	-	-	4.15	27.5	-	100	100	100	-	-	$HK^2_{PP} + Hpt^1 \rightarrow HK^2_{P0} + Hpt^1_P$
	-	-	-	-	-	10	-	-	-	-	100	$HK^2_{PP} + Hpt^2 \rightarrow HK^2_{P0} + Hpt^2_P$
$k_{2,7}$	-	-	-	-	-	-	100	10	10	-	-	$HK^2_{00} + Hpt^1_P \rightarrow HK^2_{0P} + Hpt^1$
$k_{2,8}$	-	-	-	-	-	-	10	10	10	-	-	$HK^2_{P0} + Hpt^1_P \rightarrow HK^2_{PP} + Hpt^1$
$k_{2,9}$	-	-	-	-	-	-	10	10	10	-	-	$HK^2_{0P} \rightarrow HK^2_{00}$
$k_{2,10}$	-	-	-	-	-	-	10	10	10	-	-	$HK^2_{PP} \rightarrow HK^2_{P0}$
$k_{2,11}$	-	-	-	-	-	0.1	-	-	-	-	2	$Hpt^2_P + RR \rightarrow Hpt^2 + RR_P$
$k_{3,1}$	-	-	-	-	0.0658	-	-	-	-	-	$k_{1,1}$	$HK^3_{00} \rightarrow HK^3_{P0}$
$k_{3,2}$	-	-	-	-	6000	-	-	-	-	-	100	$HK^3_{P0} \rightarrow HK^3_{0P}$
$k_{3,3}$	-	-	-	-	3.442	-	-	-	-	-	$10 \times k_{1,1}$	$HK^3_{0P} \rightarrow HK^3_{PP}$
$k_{3,4}$	-	-	-	-	0.658	-	-	-	-	-	-	$HK^3_{0P} + Hpt^1 \rightarrow HK^3_{00} + Hpt^1_P$
	-	-	-	-	-	-	-	-	-	-	50	$HK^3_{0P} + Hpt^3 \rightarrow HK^3_{00} + Hpt^3_P$
$k_{3,5}$	-	-	-	-	157.7	-	-	-	-	-	-	$HK^3_{PP} + Hpt^1 \rightarrow HK^3_{P0} + Hpt^1_P$
	-	-	-	-	-	-	-	-	-	-	100	$HK^3_{PP} + Hpt^3 \rightarrow HK^3_{P0} + Hpt^3_P$
$k_{3,11}$	-	-	-	-	-	-	-	-	-	-	5	$Hpt^3_P + RR \rightarrow Hpt^3 + RR_P$
k_{13}	-	-	-	-	-	10	14.6019	10	7	20	10	$RR_P \rightarrow RR$
HK^1_{tot}	1	1	7	6.38	16.93	140	0.18687	0.15	0.5	1	1	
HK^2_{tot}	-	-	-	7.84	27.58	100	0.18687	0.15	0.5	-	1	
HK^3_{tot}	-	-	-	-	152	-	-	-	-	-	1	
Hpt^1_{tot}	1	1	1	16.27	3.4	170	5	5	5	5	1	
Hpt^2_{tot}	-	-	-	-	-	90	-	-	-	-	1	
Hpt^3_{tot}	-	-	-	-	-	-	-	-	-	-	1	
RR_{tot}	-	-	-	-	-	100	1	1	1	1	1	

[a]Figure 3.2: initial concentrations are $[HK_{P0}]^{initial} = 0.031143$, $[HK_{0P}]^{initial} = 0.033794$, $[HK_{PP}]^{initial} = 0.065681$, $[Hpt_P]^{initial} = 0.3$ .. 1

[b]Figure 3.8: the mesh size of the surface grid is 0.1

Reactions with corresponding rates set to zero in all analyses were not included in the table. There might be small differences between the names for rate constants here and in chapter 3.

G
APPENDIX

G.1 List of Species used in section 4.2.1.1

Bordenstein Species Found in P2CS	Life Style	Bordenstein Species Found in P2CS	Life Style	Bordenstein Species Found in P2CS	Life Style	Bordenstein Species Found in P2CS	Life Style
AcidithiobacillusferrooxidansATCC23270	Extracellular	*Lactobacillusdelbrueckiisubsp.bulgaricusATCCBAA-365*	Extracellular	*Brucellasuis1330*	Facultative	*PhotobacteriumprofundumSS9*	Extracellular
Acinetobactersp.ADP1	Extracellular	*LactobacillusjohnsoniiNCC533*	Extracellular	*BurkholderiacenocepaciaAU1054*	Facultative	*Polaromonassp.JS666*	Extracellular
AlcanivoraxborkumensisSK2	Extracellular	*LactobacillusplantarumWCFS1*	Extracellular	*BurkholderiacenocepaciaAU1054*	Facultative	*PorphyromonasgingivalisW83*	Extracellular
Anaeromyxobacterdehalogenans2CP-C	Extracellular	*LawsoniaintracellularisPHE/MN1-00*	Obligate	*BurkholderiacenocepaciaAU1054*	Facultative	*Prevotellaintermedia17*	Extracellular
AquifexaeolicusVF5	Extracellular	*Leuconostocmesenteroidessubsp.mesenteroidesATCC8293*	Extracellular	*BurkholderiacenocepaciaHI2424*	Facultative	*PropionibacteriumacnesKPA171202*	Extracellular
BacilluscereusATCC14579	Extracellular	*ListeriamonocytogenesEGD-e*	Facultative	*BurkholderiacenocepaciaHI2424*	Facultative	*PseudoalteromonasatlanticaT6c*	Extracellular
BacillusclausiiKSM-K16	Extracellular	*Listeriawelshimeriserovar6bstr.SLCC5334*	Facultative	*BurkholderiacenocepaciaHI2424*	Facultative	*PseudoalteromonashaloplanktisTAC125*	Extracellular
BacillushaloduransC-125	Extracellular	*MagnetospirillummagneticumAMB-1*	Extracellular	*BurkholderiamalleiATCC23344*	Facultative	*PseudoalteromonashaloplanktisTAC125*	Extracellular
BacilluslicheniformisATCC14580(DSM13)	Extracellular	*MannheimiasucciniciproducensMBEL55E*	Extracellular	*BurkholderiamalleiATCC23344*	Facultative	*PseudomonasaeruginosaPAO1*	Extracellular
BacilluslicheniformisATCC14580(DSM13)	Extracellular	*MesorhizobiumlotiMAFF303099*	Facultative	*BurkholderiamalleiNCTC10229*	Facultative	*PseudomonasentomophilaL48*	Extracellular
BacteroidesfragilisNCTC9343	Extracellular	*MethylibiumpetroleiphilumPM1*	Extracellular	*BurkholderiamalleiNCTC10229*	Facultative	*PseudomonasfluorescensPf-5*	Extracellular
BacteroidesfragilisYCH46	Extracellular	*MethylobacillusflagellatusKT*	Extracellular	*BurkholderiamalleiNCTC10247*	Facultative	*PseudomonasputidaKT2440*	Extracellular
BacteroidesthetaiotaomicronVPI-5482	Extracellular	*MoorellathermoaceticaATCC39073*	Extracellular	*BurkholderiamalleiNCTC10247*	Facultative	*PsychrobactercryohalolentisK5*	Extracellular
BartonellabacilliformisKC583	Facultative	*Mycobacteriumavium104*	Facultative	*BurkholderiamalleiSAVP1*	Facultative	*RalstoniaeutrophaJMP134*	Extracellular
BdellovibriobacteriovorusHD100	Extracellular	*MycobacteriumlepraeTN*	Obligate	*Burkholderiapseudomallei1710b*	Facultative	*RalstoniaeutrophaJMP134*	Extracellular
BifidobacteriumlongumNCC2705	Extracellular	*MycobacteriumtuberculosisCDC1551*	Facultative	*Burkholderiapseudomallei1710b*	Facultative	*RalstoniasolanacearumGMI1000*	Extracellular
BordetellabronchisepticaRB50	Extracellular	*MyxococcusxanthusDK1622*	Extracellular	*Burkholderiapseudomallei668*	Facultative	*RenibacteriumsalmoninarumATCC33209*	Facultative
Bordetellaparapertussis12822	Extracellular	*NeisseriameningitidisMC58*	Extracellular	*Burkholderiapseudomallei668*	Facultative	*RhizobiumetliCFN42*	Facultative
BordetellapertussisTohamaI	Extracellular	*NeisseriameningitidisZ2491*	Extracellular	*BurkholderiapseudomalleiK96243*	Facultative	*RhodopseudomonaspalustrisBisB18*	Extracellular
BorreliaafzeliiPKo	Extracellular	*NitrobacterhamburgensisX14*	Extracellular	*BurkholderiapseudomalleiK96243*	Facultative	*RhodopseudomonaspalustrisBisB5*	Extracellular
BorreliaafzeliiPKo	Extracellular	*NitrobacterwinogradskyiNb-255*	Extracellular	*Burkholderiasp.383*	Facultative	*RhodopseudomonaspalustrisCGA009*	Extracellular
BorreliaburgdorferiB31	Extracellular	*NitrosococcusoceaniATCC19707*	Extracellular	*Burkholderiasp.383*	Facultative	*RhodopseudomonaspalustrisHaA2*	Extracellular
BorreliagariniiPBi	Extracellular	*NitrosomonaseuropaeaATCC19718*	Extracellular	*Burkholderiasp.383*	Facultative	*RhodospirillumrubrumATCC11170*	Extracellular
BradyrhizobiumjaponicumUSDA110	Facultative	*NitrosospiramultiformisATCC25196*	Extracellular	*BurkholderiathailandensisE264*	Facultative	*RoseobacterdenitrificansOCh114*	Extracellular
BrucellaovisATCC25840	Facultative	*NocardiafarcinicaIFM10152*	Facultative	*BurkholderiathailandensisE264*	Facultative	*RubrobacterxylanophilusDSM9941*	Extracellular
BrucellaovisATCC25840	Facultative	*NovosphingobiumaromaticivoransDSM12444*	Extracellular	*BurkholderiavietnamiensisG4*	Facultative	*Saccharophagusdegradans2-40*	Extracellular
Brucellasuis1330	Facultative	*OceanobacillusiheyensisHTE831*	Extracellular	*BurkholderiavietnamiensisG4*	Facultative	*SalinibacterruberDSM13855*	Extracellular
Brucellasuis1330	Facultative	*PelobactercarbinolicusDSM2380*	Extracellular	*BurkholderiavietnamiensisG4*	Facultative	*ShewanelladenitrificansOS217*	Extracellular
Brucellasuis1330	Facultative	*PhotobacteriumprofundumSS9*	Extracellular	*BurkholderiaxenovoransLB400*	Facultative	*ShewanellaoneidensisMR-1*	Extracellular
BurkholderiaxenovoransLB400	Facultative	*Campylobacterconcisus13826*	Extracellular	*BurkholderiaxenovoransLB400*	Facultative	*Shewanellasp.MR-4*	Extracellular
Shewanellasp.MR-7	Extracellular	*ShigellaboydiiSb227*	Facultative				

Bordenstein Species Found in P2CS	Life Style	*Bordenstein Species Found in P2CS*	Life Style	*Bordenstein Species Found ain P2CS*	Life Style	*Bordenstein Species Found in P2CS*	Life Style
Campylobactercurvus525.92	Extracellular	*ShigelladysenteriaeSd197*	Facultative	*Dehalococcoidesethenogenes195*	Extracellular	*Synechococcusspp.CC9311*	Extracellular
Campylobacterfetussubsp.fetus82-40	Extracellular	*Shigellaflexneri2astr.301*	Facultative	*Dehalococcoidesspp.CBDB1*	Extracellular	*Synechococcusspp.CC9605*	Extracellular
CampylobacterhominisATCCBAA-381	Extracellular	*Shigellaflexneri5str.8401*	Facultative	*Deinococcusgeothermalis DSM11300*	Extracellular	*Synechococcusspp.WH8102*	Extracellular
CampylobacterjejuniRM1221	Facultative	*ShigellasonneiSs046*	Facultative	*DeinococcusradioduransR1*	Extracellular	*SyntrophusaciditrophicusSB*	Extracellular
Campylobacterjejunisubsp.doylei269.97	Facultative	*Sinorhizobiummeliloti1021*	Facultative	*DeinococcusradioduransR1*	Extracellular	*ThermobifidafuscaYX*	Extracellular
Campylobacterjejunisubsp.jejuni81-176	Facultative	*Sodalisglossinidiusstr.morsitans*	Facultative	*Desulfitobacteriumhafniense Y51*	Extracellular	*ThermotogamaritimaMSB8*	Extracellular
CandidatusPelagibacterubiqueHTCC1062	Extracellular	*SphingopyxisalaskensisRB2256*	Extracellular	*Desulfotaleapsychrophila LSv54*	Extracellular	*ThermusthermophilusHB27*	Extracellular
CandidatusProtochlamydiaamoebophilaUWE25	Obligate	*StaphylococcusaureusRF122*	Extracellular	*Enterococcusfaecalis V583*	Extracellular	*ThermusthermophilusHB8*	Extracellular
Carboxydothermushydrogenoformans Z-2901	Extracellular	*Staphylococcusaureussubsp.aureusCOL*	Extracellular	*Escherichiacoli536*	Extracellular	*ThiobacillusdenitrificansATCC25259*	Extracellular
CaulobactercrescentusCB15	Extracellular	*Staphylococcusaureussubsp.aureusMRSA252*	Extracellular	*EscherichiacoliCFT073*	Extracellular	*ThiomicrospiracrunogenaXCL-2*	Extracellular
ChlamydiatrachomatisA/HAR-13	Obligate	*Staphylococcusaureussubsp.aureusMSSA476*	Extracellular	*EscherichiacoliSMS-3-5*	Extracellular	*TreponemadenticolaATCC35405*	Extracellular
ChlorobiumchlorochromatiiCaD3	Extracellular	*StaphylococcusepidermidisATCC12228*	Extracellular	*EscherichiacoliUTI89*	Extracellular	*TrichodesmiumerythraeumIMS101*	Extracellular
ChlorobiumtepidumTLS	Extracellular	*StaphylococcusepidermidisRP62A*	Extracellular	*FrankiaalniACN14a*	Extracellular	*TropherymawhippleiTW08/27*	Obligate
ChromobacteriumviolaceumATCC12472	Extracellular	*StaphylococcushaemolyticusJCSC1435*	Extracellular	*Frankiasp.CcI3*	Extracellular	*Tropherymawhippleistr.Twist*	Obligate
ChromohalobactersalexigensDSM3043	Extracellular	*Streptococcusagalactiae2603V/R*	Extracellular	*GeobacilluskaustophilusHTA426*	Extracellular	*VibriocholeraeO395*	Extracellular
ClostridiumacetobutylicumATCC824	Extracellular	*StreptococcusagalactiaeA909*	Extracellular	*Geobactermetallireducens GS-15*	Extracellular	*VibriocholeraeO395*	Extracellular
ClostridiumbeijerinckiiNCIMB8052	Extracellular	*StreptococcusagalactiaeNEM316*	Extracellular	*GeobactersulfurreducensPCA*	Extracellular	*VibriocholeraeO395*	Extracellular
ClostridiumnovyiNT	Extracellular	*StreptococcusmutansUA159*	Extracellular	*GlocobacterviolaceusPCC7421*	Extracellular	*VibriofischeriES114*	Extracellular
ClostridiumperfringensATCC13124	Extracellular	*Streptococcuspneumoniae R6*	Extracellular	*Gluconobacteroxydans621H*	Extracellular	*VibriofischeriES114*	Extracellular
ClostridiumperfringensSM101	Extracellular	*StreptococcuspneumoniaeTIGR4*	Extracellular	*GranulibacterbethesdensisCGDNIH1*	Extracellular	*VibrioparahaemolyticusRIMD2210633*	Extracellular
ClostridiumphytofermentansISDg	Extracellular	*StreptococcuspyogenesMGAS10394*	Extracellular	*Haemophilusducreyi35000HP*	Extracellular	*VibrioparahaemolyticusRIMD2210633*	Extracellular
ClostridiumtetaniE88	Extracellular	*StreptococcuspyogenesMGAS315*	Extracellular	*HahellachejuensisKCTC2396*	Extracellular	*VibriovulnificusCMCP6*	Extracellular
Colwelliapsychrerythraea34H	Extracellular	*StreptococcuspyogenesMGAS8232*	Extracellular	*Helicobacteracinonychisstr.Sheeba*	Extracellular	*VibriovulnificusCMCP6*	Extracellular
CorynebacteriumdiphtheriaeNCTC13129	Extracellular	*StreptococcuspyogenesSSI-1*	Extracellular	*HelicobacterhepaticusATCC51449*	Extracellular	*VibriovulnificusYJ016*	Extracellular
CorynebacteriumefficiensYS-314	Extracellular	*StreptococcusthermophilusCNRZ1066*	Extracellular	*Helicobacterpylori26695*	Extracellular	*VibriovulnificusYJ016*	Extracellular
CorynebacteriumjeikeiumK411	Extracellular	*StreptococcusthermophilusLMG18311*	Extracellular	*HelicobacterpyloriHPAG1*	Extracellular	*Xanthomonasoryzaepv.oryzaeMAFF311018*	Extracellular
CoxiellaburnetiiRSA331	Obligate	*Streptomycesavermitilis MA-4680*	Extracellular	*HelicobacterpyloriJ99*	Extracellular	*Xylellafastidiosa9a5c*	Extracellular
CoxiellaburnetiiRSA493	Obligate	*Streptomycescoelicolor A3(2)*	Extracellular	*HyphomonasneptuniumATCC15444*	Extracellular	*XylellafastidiosaTemecula1*	Extracellular
CytophagahutchinsoniiATCC33406	Extracellular	*SymbiobacteriumthermophilumIAM14863*	Extracellular	*IdiomarinaloihiensisL2TR*	Extracellular	*YersiniapestisAngola*	Facultative
DechloromonasaromaticaRCB	Extracellular	*Synechococcuselongatus PCC7942*	Extracellular	*Jannaschiasp.CCS1*	Extracellular	*YersiniapestisCO92*	Facultative
LactobacillusbrevisATCC367	Extracellular	*YersiniapseudotuberculosisIP31758*	Facultative	*LactobacillusacidophilusNCFM*	Extracellular	*YersiniapestisNepal516*	Facultative
Lactobacillusdelbrueckiisubsp.bulgaricusATCC11842	Extracellular	*YersiniapseudotuberculosisIP32953*	Facultative				

Bibliography

[1] Agnes Rodrigue, Yves Quentin, Andree Lazdunski, Vincent Mejean, and Maryline Foglino. Cell signalling by oligosaccharides. two-component systems in pseudomonas aeruginosa: why so many? *Trends in Microbiology*, 8 (11):498–504, 2016/03/22 2000. doi: 10.1016/S0966-842X(00)01833-3. URL `http://dx.doi.org/10.1016/S0966-842X(00)01833-3`.

[2] Franziska Mika and Regine Hengge. A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of ÏČS (RpoS) in E. coli. *Genes & Development*, 19(22):2770–2781, November 2005. ISSN 0890-9369, 1549-5477. doi: 10.1101/gad.353705. URL `http://genesdev.cshlp.org/content/19/22/2770`.

[3] Ilka B Bischofs, Joshua a Hug, Aiwen W Liu, Denise M Wolf, and Adam P Arkin. Complexity in bacterial cell-cell communication: quorum signal integration and subpopulation signaling in the Bacillus subtilis phosphorelay. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6459–64, April 2009. ISSN 1091-6490. doi: 10.1073/pnas.0810878106. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2672556&tool=pmcentrez&rendertype=abstract`.

[4] Kenny C Mok, Ned S Wingreen, and Bonnie L Bassler. Vibrio harveyi quorum sensing: a coincidence detector for two autoinducers controls gene expression. *The EMBO journal*, 22(4):870–881, February 2003. ISSN 0261-4189. doi: 10.1093/emboj/cdg085.

[5] Akinori Kato, Alexander Y. Mitrophanov, and Eduardo A. Groisman. A connector of two-component regulatory systems promotes signal amplification and persistence of expression. *Proceedings of the National Academy of Sciences*, 104(29):12063–12068, 2007. doi: 10.1073/pnas.0704462104. URL `http://www.pnas.org/content/104/29/12063.abstract`.

[6] Michael A. Rowland and Eric J. Deeds. Crosstalk and the evolution of specificity in two-component signaling. *Proceedings of the National Academy of Sciences*,

111(15):5550–5555, 2014. doi: 10.1073/pnas.1317178111. URL http://www.pnas.org/content/early/2014/03/27/1317178111.

[7] Michael T Laub and Mark Goulian. Specificity in two-component signal transduction pathways. *Annual review of genetics*, 41:121–45, January 2007. ISSN 0066-4197. doi: 10.1146/annurev.genet.41.042007.170548. URL http://www.ncbi.nlm.nih.gov/pubmed/18076326.

[8] U. S. Bhalla and R. Iyengar. Emergent properties of networks of biological signaling pathways. *Science (New York, N.Y.)*, 283(5400):381–387, January 1999. ISSN 0036-8075.

[9] Albert Siryaporn and Mark Goulian. Cross-talk suppression between the CpxA-CpxR and EnvZ-OmpR two-component systems in E. coli. *Molecular microbiology*, 70(2):494–506, October 2008. ISSN 1365-2958. doi: 10.1111/j.1365-2958.2008.06426.x.

[10] Joseph J. Falke, Randal B. Bass, Scott L. Butler, Stephen A. Chervitz, and Mark A. Danielson. The two-component signaling pathway of bacterial chemotaxis: A molecular view of signal transduction by receptors, kinases, and adaptation enzymes the two-component signaling pathway of bacterial chemotaxis: A molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annual review of cell and developmental biology*, 13:457–512, 1997. ISSN 1081-0706. doi: 10.1146/annurev.cellbio.13.1.457. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2899694/.

[11] Ann M. Stock, Victoria L. Robinson, and Paul N. Goudreau. Two-Component Signal Transduction. *Annual Review of Biochemistry*, 69(1):183–215, 2000. doi: 10.1146/annurev.biochem.69.1.183. URL http://dx.doi.org/10.1146/annurev.biochem.69.1.183.

[12] Michael Y Galperin. Structural classification of bacterial response regulators: diversity of output domains and domain combinations. *Journal of bacteriology*, 188(12):4169–4182, June 2006. ISSN 0021-9193. doi: 10.1128/JB.01887-05.

[13] Michael Y. Galperin. Diversity of Structure and Function of Response Regulator Output Domains. *Current opinion in microbiology*, 13(2):150–159, April 2010. ISSN 1369-5274. doi: 10.1016/j.mib.2010.01.005. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3086695/.

[14] Francesc Posas, Susannah M Wurgler-Murphy, Tatsuya Maeda, Elizabeth A Witten, Tran Cam Thai, and Haruo Saito. Yeast HOG1 MAP Kinase Cascade

Is Regulated by a Multistep Phosphorelay Mechanism in the SLN1–YPD1–SSK1 "Two-Component" Osmosensor. *Cell*, 86(6):865–875, September 1996. ISSN 0092-8674. doi: 10.1016/S0092-8674(00)80162-2. URL http://www.sciencedirect.com/science/article/pii/S0092867400801622.

[15] Darío Ortiz de Orué Lucana and Matthew R. Groves. The three-component signalling system HbpS–SenS–SenR as an example of a redox sensing pathway in bacteria. *Amino Acids*, 37(3):479–486, September 2009. ISSN 0939-4451, 1438-2199. doi: 10.1007/s00726-009-0260-9. URL http://link.springer.com/article/10.1007/s00726-009-0260-9.

[16] Munia Amin, Varun B. Kothamachu, Elisenda Feliu, Birgit E. Scharf, Steven L. Porter, and Orkun S. Soyer. Phosphate Sink Containing Two-Component Signaling Systems as Tunable Threshold Devices. *PLoS Comput Biol*, 10 (10):e1003890, October 2014. doi: 10.1371/journal.pcbi.1003890. URL http://dx.doi.org/10.1371/journal.pcbi.1003890.

[17] M Kravanja, R Engelmann, V Dossonnet, M Blüggel, H E Meyer, R Frank, A Galinier, J Deutscher, N Schnell, and W Hengstenberg. The hprK gene of Enterococcus faecalis encodes a novel bifunctional enzyme: the HPr kinase/phosphatase. *Molecular microbiology*, 31(1):59–66, January 1999. ISSN 0950-382X.

[18] Helena Ramström, Sarah Sanglier, Emmanuelle Leize-Wagner, Claude Philippe, Alain Van Dorsselaer, and Jacques Haiech. Properties and regulation of the bifunctional enzyme HPr kinase/phosphatase in Bacillus subtilis. *The Journal of Biological Chemistry*, 278(2):1174–1185, January 2003. ISSN 0021-9258. doi: 10.1074/jbc.M209052200.

[19] Valérie Dossonnet, Vicente Monedero, Monique Zagorec, Anne Galinier, Gaspar Pérez-Martínez, and Josef Deutscher. Phosphorylation of HPr by the Bifunctional HPr Kinase/P-Ser-HPr Phosphatase from Lactobacillus casei Controls Catabolite Repression and Inducer Exclusion but Not Inducer Expulsion. *Journal of Bacteriology*, 182(9):2582–2590, May 2000. ISSN 0021-9193, 1098-5530. doi: 10.1128/JB.182.9.2582-2590.2000. URL http://jb.asm.org/content/182/9/2582.

[20] Yan Zhu, Ling Qin, Takeshi Yoshida, and Masayori Inouye. Phosphatase Activity of Histidine Kinase EnvZ Without Kinase Catalytic Domain. *Proceedings of the National Academy of Sciences*, 97(14):7808–7813, July 2000. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.97.14.7808. URL http://www.pnas.org/content/97/14/7808.

[21] Joseph P Dexter and Jeremy Gunawardena. Dimerization and bifunctionality confer robustness to the isocitrate dehydrogenase regulatory system in Escherichia coli. *The Journal of biological chemistry*, 288(8):5770–5778, February 2013. ISSN 1083-351X. doi: 10.1074/jbc.M112.339226.

[22] Steven L Porter, Mark a J Roberts, Cerys S Manning, and Judith P Armitage. A bifunctional kinase-phosphatase in bacterial chemotaxis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(47):18531–6, November 2008. ISSN 1091-6490. doi: 10.1073/pnas.0808010105. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2587623&tool=pmcentrez&rendertype=abstract`.

[23] Verena Weiss, Günter Kramer, Thomas Dünnebier, and Annette Flotho. Mechanism of regulation of the bifunctional histidine kinase NtrB in Escherichia coli. *Journal of molecular microbiology and biotechnology*, 4(3):229–233, May 2002. ISSN 1464-1801.

[24] Eli S Groban, Elizabeth J Clarke, Howard M Salis, Susan M Miller, and Christopher a Voigt. Kinetic buffering of cross talk between bacterial two-component sensors. *Journal of molecular biology*, 390 (3):380–93, July 2009. ISSN 1089-8638. doi: 10.1016/j.jmb.2009.05.007. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2974629&tool=pmcentrez&rendertype=abstract`.

[25] Y. Erin Chen, Christos G. Tsokos, Emanuele G. Biondi, Barrett S. Perchuk, and Michael T. Laub. Dynamics of Two Phosphorelays Controlling Cell Cycle Progression in Caulobacter crescentus. *Journal of Bacteriology*, 191(24):7417–7429, December 2009. ISSN 0021-9193, 1098-5530. doi: 10.1128/JB.00992-09. URL `http://jb.asm.org/content/191/24/7417`.

[26] Ronny Straube. Reciprocal Regulation as a Source of Ultrasensitivity in Two-Component Systems with a Bifunctional Sensor Kinase. *PLoS Comput Biol*, 10(5):e1003614, May 2014. doi: 10.1371/journal.pcbi.1003614. URL `http://dx.doi.org/10.1371/journal.pcbi.1003614`.

[27] John S. Parkinson and Eric C. Kofoid. Communication modules in bacterial signaling proteins. *Annual Review of Genetics*, 26(1):71–112, 1992. doi: 10.1146/annurev.ge.26.120192.000443. URL `http://dx.doi.org/10.1146/annurev.ge.26.120192.000443`.

[28] Takeshi Urao, Kazuko Yamaguchi-Shinozaki, and Kazuo Shinozaki. Two-component systems in plant signal transduction. *Trends in Plant Science*,

5(2):67–74, February 2000. ISSN 1360-1385. doi: 10.1016/S1360-1385(99)01542-3. URL http://www.sciencedirect.com/science/article/pii/S1360138599015423.

[29] Jeryl L Appleby, John S Parkinson, and Robert B Bourret. Signal Transduction via the Multi-Step Phosphorelay: Not Necessarily a Road Less Traveled. *Cell*, 86(6):845–848, September 1996. ISSN 0092-8674. doi: 10.1016/S0092-8674(00)80158-0. URL http://www.sciencedirect.com/science/article/pii/S0092867400801580.

[30] Haruo Saito. Histidine phosphorylation and two-component signaling in eukaryotic cells. *Chemical Reviews*, 101(8):2497–2510, 2001. doi: 10.1021/cr000243+. URL http://dx.doi.org/10.1021/cr000243+.

[31] T Mizuno. His-Asp phosphotransfer signal transduction. *Journal of biochemistry*, 123(4):555–563, April 1998. ISSN 0021-924X.

[32] James A Hoch. Two-component and phosphorelay signal transduction. *Current Opinion in Microbiology*, 3(2):165–170, April 2000. ISSN 1369-5274. doi: 10.1016/S1369-5274(00)00070-9. URL http://www.sciencedirect.com/science/article/pii/S1369527400000709.

[33] Attila Csikász-Nagy, Luca Cardelli, and Orkun S Soyer. Response dynamics of phosphorelays suggest their potential utility in cell signalling. *Journal of the Royal Society, Interface*, 8(57):480–8, April 2011. ISSN 1742-5662. doi: 10.1098/rsif.2010.0336. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3061117&tool=pmcentrez&rendertype=abstract.

[34] Jan-Willem Veening, Leendert W Hamoen, and Oscar P Kuipers. Phosphatases modulate the bistable sporulation gene expression pattern in Bacillus subtilis. *Molecular microbiology*, 56(6):1481–1494, June 2005. ISSN 0950-382X. doi: 10.1111/j.1365-2958.2005.04659.x.

[35] Abhinav Tiwari, J Christian J Ray, Jatin Narula, and Oleg a Igoshin. Bistable responses in bacterial genetic networks: designs and dynamical consequences. *Mathematical biosciences*, 231(1):76–89, May 2011. ISSN 1879-3134. doi: 10.1016/j.mbs.2011.03.004. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3095517&tool=pmcentrez&rendertype=abstract.

[36] Arnaud Chastanet, Dennis Vitkup, Guo-Cheng Yuan, Thomas M. Norman, Jun S. Liu, and Richard M. Losick. Broadly heterogeneous activation of the master regulator for sporulation in Bacillus subtilis. *Proceedings of the National*

Academy of Sciences, 107(18):8486–8491, May 2010. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1002499107. URL http://www.pnas.org/content/107/18/8486.

[37] Jatin Narula, Seram N. Devi, Masaya Fujita, and Oleg A. Igoshin. Ultrasensitivity of the bacillus subtilis sporulation decision. *Proceedings of the National Academy of Sciences*, 109(50):E3513–E3522, 2012. doi: 10.1073/pnas.1213974109. URL http://www.pnas.org/content/109/50/E3513.abstract.

[38] M B Elowitz and S Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–8, January 2000. ISSN 0028-0836. doi: 10.1038/35002125. URL http://www.ncbi.nlm.nih.gov/pubmed/10659856.

[39] John J Tyson, Katherine C Chen, and Bela Novak. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Current Opinion in Cell Biology*, 15(2):221–231, April 2003. ISSN 0955-0674. doi: 10.1016/S0955-0674(03)00017-6. URL http://www.sciencedirect.com/science/article/pii/S0955067403000176.

[40] A Goldbeter and D E Koshland. An amplified sensitivity arising from covalent modification in biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, 78(11):6840–6844, November 1981. ISSN 0027-8424. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC349147/.

[41] Peter A. Spiro, John S. Parkinson, and Hans G. Othmer. A model of excitation and adaptation in bacterialâĂĽchemotaxis. *Proceedings of the National Academy of Sciences*, 94(14):7263–7268, 1997. URL http://www.pnas.org/content/94/14/7263.abstract.

[42] Ann H West and Ann M Stock. Histidine kinases and response regulator proteins in two-component signaling systems. *Trends in Biochemical Sciences*, 26(6):369–376, June 2001. ISSN 0968-0004. doi: 10.1016/S0968-0004(01)01852-7. URL http://www.sciencedirect.com/science/article/pii/S0968000401018527.

[43] V L Robinson, D R Buckler, and a M Stock. A tale of two components: a novel kinase and a regulatory switch. *Nature structural biology*, 7(8):626–633, August 2000. ISSN 1072-8368. doi: 10.1038/77915.

[44] John C Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, http://onlinelibrary.wiley.com/book/10.1002/9780470753767, 2nd edition edition, March 2003.

[45] F. Horn and R. Jackson. General mass action kinetics. *Archive for Rational Mechanics and Analysis*, 47(2):81–116, 1972. ISSN 1432-0673. doi: 10.1007/BF00251225. URL http://dx.doi.org/10.1007/BF00251225.

[46] I. M. Griffiths, D. V.; Smith. *Numerical methods for engineers: a programming approach.* Number ISBN 0-8493-8610-1 in -. CRC Press, Boca Raton, FL, USA, 1st edition edition, 1991.

[47] W. H.; Flannery B. P.; Teukolsky S. A.; Press and Vetterling W. T. *Numerical Recipes in FORTRAN: The Art of Scientific Computing.* Number p. 710 in -. Cambridge University Press, Cambridge, England, 2nd edition edition, 1992.

[48] G. Arfken. *Mathematical Methods for Physicists, 3rd ed.* Academic Press, Orlando, FL, 1985.

[49] MATLAB. *version 7.11.0 (R2010b).* The MathWorks Inc., Natick, Massachusetts, 2010.

[50] Shampine L. F. and M. W. Reichelt. The matlab ode suite. *SIAM Journal on Scientific Computing,*, Vol. 18:1–22, 1997.

[51] Lawrence F. Shampine, Mark W. Reichelt, and Jacek A. Kierzenka. Solving index-1 daes in matlab and simulink. *SIAM Rev.*, 41(3):538–552, September 1999. ISSN 0036-1445. doi: 10.1137/S003614459933425X. URL http://dx.doi.org/10.1137/S003614459933425X.

[52] Bard Ermentrout. *XPPAUT.* Springer Netherlands, 2012. ISBN 978-94-007-3858-4. doi: 10.1007/978-94-007-3858-417. URL http://dx.doi.org/10.1007/978-94-007-3858-4_17.

[53] Vijay Chickarmane, Sri R Paladugu, Frank Bergmann, and Herbert M Sauro. Bifurcation discovery tool. *Bioinformatics (Oxford, England)*, 21(18):3688–90, September 2005. ISSN 1367-4803. doi: 10.1093/bioinformatics/bti603. URL http://www.ncbi.nlm.nih.gov/pubmed/16081475.

[54] Marcus J. Tindall, Steven L. Porter, Philip K. Maini, and Judith P. Armitage. Modeling Chemotaxis Reveals the Role of Reversed Phosphotransfer and a Bi-Functional Kinase-Phosphatase. *PLoS Computational Biology*, 6(8):e1000896, August 2010. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000896. URL http://dx.plos.org/10.1371/journal.pcbi.1000896.

[55] C E Grimshaw, S Huang, C G Hanstein, M a Strauch, D Burbulys, L Wang, J a Hoch, and J M Whiteley. Synergistic kinetic interactions between components of the phosphorelay controlling sporulation in Bacillus subtilis. *Biochemistry*,

37(5):1365–75, February 1998. ISSN 0006-2960. doi: 10.1021/bi971917m. URL http://www.ncbi.nlm.nih.gov/pubmed/9477965.

[56] Strogatz S. *Nonlinear Dynamics and Chaos: with applications to physics, biology, chemistry, and engineering.* Westview Press., 1994.

[57] Laura A. Mike, Jacob E. Choby, Paul R. Brinkman, Lorenzo Q. Olive, Brendan F. Dutter, Samuel J. Ivan, Christopher M. Gibbs, Gary A. Sulikowski, Devin L. Stauff, and Eric P. Skaar. Two-Component System Cross-Regulation Integrates Bacillus anthracis Response to Heme and Cell Envelope Stress. *PLoS Pathog*, 10(3):e1004044, March 2014. doi: 10.1371/journal.ppat.1004044. URL http://dx.doi.org/10.1371/journal.ppat.1004044.

[58] Katy Wei, Maxim Moinat, Timo R. Maarleveld, and Frank J. Bruggeman. Stochastic simulation of prokaryotic two-component signalling indicates stochasticity-induced active-state locking and growth-rate dependent bistability. *Molecular bioSystems*, 10(9):2338–2346, July 2014. ISSN 1742-2051. doi: 10.1039/c4mb00264d.

[59] Xiao Wang, Nan Hao, Henrik G. Dohlman, and Timothy C. Elston. Bistability, Stochasticity, and Oscillations in the Mitogen-Activated Protein Kinase Cascade. *Biophysical Journal*, 90(6):1961–1978, March 2006. ISSN 0006-3495. doi: 10.1529/biophysj.105.073874. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1386776/.

[60] Jan-Willem Veening, Wiep Klaas Smits, and Oscar P. Kuipers. Bistability, Epigenetics, and Bet-Hedging in Bacteria. *Annual Review of Microbiology*, 62(1):193–210, 2008. doi: 10.1146/annurev.micro.62.081307.163002. URL http://www.annualreviews.org/doi/abs/10.1146/annurev.micro.62.081307.163002.

[61] J. Saez-Rodriguez, A. Hammerle-Fickinger, O. Dalal, S. Klamt, E. D. Gilles, and C. Conradi. Multistability of signal transduction motifs. *IET Systems Biology*, 2(2):80–93, 2008. ISSN 1751-8849. doi: 10.1049/iet-syb:20070012.

[62] Dan Siegal-Gaskins, Erich Grotewold, and Gregory D. Smith. The capacity for multistability in small gene regulatory networks. *BMC Systems Biology*, 3 (1):96, September 2009. ISSN 1752-0509. doi: 10.1186/1752-0509-3-96. URL http://www.biomedcentral.com/1752-0509/3/96/abstract.

[63] R. Thomas. Numerical methods in the study of critical phenomena. In Jean Della Dora, Jacques Demongeot, and Bernard Lacolle, editors, *Proceedings of a Colloquium, Carry-le-Rouet*, pages 180–193, Berlin, Heidel-

berg, June 1980. Springer Berlin Heidelberg. ISBN 978-3-642-81703-8. doi: 10.1007/978-3-642-81703-8_24. URL http://dx.doi.org/10.1007/978-3-642-81703-8_24.

[64] C. Soulé. Graphic requirements for multistationarity. *Complexus*, 1(3):123–133, 2003. URL http://www.karger.com/DOI/10.1159/000076100.

[65] M. Kaufman, C. Soulé, and R. Thomas. A new necessary condition on interaction graphs for multistationarity. *Journal of Theoretical Biology*, 248 (4):675 – 685, 2007. ISSN 0022-5193. doi: http://dx.doi.org/10.1016/j.jtbi.2007.06.016. URL http://www.sciencedirect.com/science/article/pii/S0022519307002950.

[66] DeLisi C Eisenfeld J. *On conditions for qualitative instability of regulatory circuits with application to immunological control loops.* Number pg 39-53 in Mathematics and Computers in Biomedical Applications. Elsevier Science Publishers, 1985.

[67] Martin Feinberg. Chemical oscillations, multiple equilibria, and reaction network structure. In Ray W Harmon Stewart Warren E and Conley Charles C, editors, *Dynamics and Modelling of Reactive Systems*, pages 59 – 130. Academic Press, 1980. ISBN 978-0-12-669550-2. doi: http://dx.doi.org/10.1016/B978-0-12-669550-2.50008-7. URL http://www.sciencedirect.com/science/article/pii/B9780126695502500087.

[68] Elisenda Feliu. Injectivity, multiple zeros and multistationarity in reaction networks. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 471(2173), 12 2014.

[69] Gheorghe Craciun and Martin Feinberg. Multiple equilibria in complex chemical reaction networks: I. the injectivity property. *SIAM Journal on Applied Mathematics*, 65(5):1526–1546, 2005. doi: 10.1137/S0036139904440278. URL http://dx.doi.org/10.1137/S0036139904440278.

[70] G. Craciun and M. Feinberg. Multiple equilibria in complex chemical reaction networks: Ii. the species-reaction graph. *SIAM Journal on Applied Mathematics*, 66(4):1321–1338, 2016/05/11 2006. doi: 10.1137/050634177. URL http://dx.doi.org/10.1137/050634177.

[71] Gheorghe Craciun, Yangzhong Tang, and Martin Feinberg. Understanding bistability in complex enzyme-driven reaction networks. *Proceedings of the*

National Academy of Sciences of the United States of America, 103(23):8697–8702, 06 2006. doi: 10.1073/pnas.0602767103. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1592242/.

[72] Paul M. Schlosser and Martin Feinberg. A theory of multiple steady states in isothermal homogeneous cfstrs with many reactions. *Chemical Engineering Science*, 49(11):1749 – 1767, 1994. ISSN 0009-2509. doi: http://dx.doi.org/10.1016/0009-2509(94)80061-8. URL http://www.sciencedirect.com/science/article/pii/0009250994800618.

[73] Martin Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors—I. The deficiency zero and deficiency one theorems. *Chemical Engineering Science*, 42(10):2229–2268, 1987. ISSN 00092509. doi: 10.1016/0009-2509(87)80099-4. URL http://dx.doi.org/10.1016/0009-2509(87)80099-4.

[74] Martin Feinberg. Chemical reaction network structure and the stability of complex isothermal reactors—ii. multiple steady states for networks of deficiency one. *Chemical Engineering Science*, 43(1):1 – 25, 1988. ISSN 0009-2509. doi: http://dx.doi.org/10.1016/0009-2509(88)87122-7. URL http://www.sciencedirect.com/science/article/pii/0009250988871227.

[75] Phillipp Ellison and Martin Feinberg. How catalytic mechanisms reveal themselves in multiple steady-state data: I. Basic principles. *Journal of Molecular Catalysis A: Chemical*, 154(1–2):155–167, March 2000. ISSN 1381-1169. doi: 10.1016/S1381-1169(99)00371-4. URL http://www.sciencedirect.com/science/article/pii/S1381116999003714.

[76] Martin Feinberg. Multiple steady states for chemical reaction networks of deficiency one. *Archive for Rational Mechanics and Analysis*, 132(4):371–406, 1995. ISSN 1432-0673. doi: 10.1007/BF00375615. URL http://dx.doi.org/10.1007/BF00375615.

[77] Martin Feinberg. The existence and uniqueness of steady states for a class of chemical reaction networks. *Archive for Rational Mechanics and Analysis*, 132(4):311–370, 1995. ISSN 1432-0673. doi: 10.1007/BF00375614. URL http://dx.doi.org/10.1007/BF00375614.

[78] Alexander Y Mitrophanov, Gordon Churchward, and Mark Borodovsky. Control of Streptococcus pyogenes virulence: modeling of the CovR/S signal transduction system. *Journal of theoretical biology*, 246 (1):113–28, May 2007. ISSN 0022-5193. doi: 10.1016/j.jtbi.2006.11.

009. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2688695&tool=pmcentrez&rendertype=abstract.

[79] Nils Blüthgen, Stefan Legewie, Hanspeter Herzel, and Boris Kholodenko. Mechanisms Generating Ultrasensitivity, Bistability, and Oscillations in Signal Transduction. In Sangdun Choi, editor, *Introduction to Systems Biology*, pages 282–299. Humana Press, January 2007. ISBN 978-1-58829-706-8, 978-1-59745-531-2. URL http://link.springer.com/chapter/10.1007/978-1-59745-531-2_15.

[80] Phillipp Ellison and Martin Feinberg. How catalytic mechanisms reveal themselves in multiple steady-state data: I. basic principles. *Journal of Molecular Catalysis A: Chemical*, 154(1–2):155–167, 3 2000. doi: http://dx.doi.org/10.1016/S1381-1169(99)00371-4. URL http://www.sciencedirect.com/science/article/pii/S1381116999003714.

[81] Agnès Rodrigue, Yves Quentin, Andrée Lazdunski, Vincent Méjean, and Maryline Foglino. Cell signalling by oligosaccharides. Two-component systems in Pseudomonas aeruginosa: why so many? *Trends in Microbiology*, 8(11):498–504, November 2000. ISSN 0966-842X. doi: 10.1016/S0966-842X(00)01833-3. URL http://www.sciencedirect.com/science/article/pii/S0966842X00018333.

[82] Fabiola Janiak-Spens, Paul F Cook, and Ann H West. Kinetic analysis of YPD1-dependent phosphotransfer reactions in the yeast osmoregulatory phosphorelay system. *Biochemistry*, 44(1):377–386, January 2005. ISSN 0006-2960. doi: 10.1021/bi048433s.

[83] L. N. Johnson and R. J. Lewis. Structural basis for control by phosphorylation. *Chemical Reviews*, 101(8):2209–2242, August 2001. ISSN 0009-2665.

[84] D Georgellis, a S Lynch, and E C Lin. In vitro phosphorylation study of the arc two-component signal transduction system of Escherichia coli. *Journal of bacteriology*, 179(17):5429–35, October 1997. ISSN 0021-9193. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=179413&tool=pmcentrez&rendertype=abstract.

[85] Varun B. Kothamachu, Elisenda Feliu, Carsten Wiuf, Luca Cardelli, and Orkun S. Soyer. Phosphorelays Provide Tunable Signal Processing Capabilities for the Cell. *PLoS Comput Biol*, 9(11):e1003322, November 2013. doi: 10.1371/journal.pcbi.1003322. URL http://dx.doi.org/10.1371/journal.pcbi.1003322.

[86] M. Kwiatkowska, G. Norman, and D. Parker. Stochastic model checking. In M. Bernardo and J. Hillston, editors, *Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation (SFM'07)*, volume 4486 of *LNCS (Tutorial Volume)*, pages 220–270. Springer, 2007.

[87] M. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of Probabilistic Real-time Systems. In G. Gopalakrishnan and S. Qadeer, editors, *Proc. 23rd International Conference on Computer Aided Verification (CAV'11)*, volume 6806 of *LNCS*, pages 585–591. Springer, 2011.

[88] M. Kwiatkowska, G. Norman, and D. Parker. Stochastic Model Checking. In M. Bernardo and J. Hillston, editors, *Formal Methods for the Design of Computer, Communication and Software Systems: Performance Evaluation (SFM'07)*, volume 4486 of *LNCS (Tutorial Volume)*, pages 220–270. Springer, 2007.

[89] Luca Cardelli and Attila Csikász-Nagy. The Cell Cycle Switch Computes Approximate Majority. *Scientific Reports*, 2, September 2012. ISSN 2045-2322. doi: 10.1038/srep00656. URL `http://www.nature.com/srep/2012/120913/srep00656/full/srep00656.html#/methods`.

[90] Luca Cardelli. From Processes to ODEs by Chemistry. *Differential Equations*, pages 1–16, 2008.

[91] L Cardelli. On process rate semantics. *Theoretical Computer Science*, 391(3): 190–215, February 2008. ISSN 03043975. doi: 10.1016/j.tcs.2007.11.012. URL `http://linkinghub.elsevier.com/retrieve/pii/S0304397507008559`.

[92] M.L. Menten L. Michaelis. Kinetik der invertinwirkung. *Biochem. Zeitung, 49 (1913), pp. 333–369*, 1913.

[93] Kenneth A Johnson and Roger S Goody. The original michaelis constant: Translation of the 1913 michaelis-menten paper. *Biochemistry*, 50(39):8264–8269, 10 2011. doi: 10.1021/bi201284u. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3381512/`.

[94] Stryer L. Berg JM, Tymoczko JL. *Biochemistry. 5th edition.* The Michaelis-Menten Model Accounts for the Kinetic Properties of Many Enzymes. New York: W H Freeman; Section 8.4, 2002.

[95] Elisenda Feliu and Carsten Wiuf. Simplifying biochemical models with intermediate species. *Journal of the Royal Society Interface*, 10(87), Oc-

tober 2013. ISSN 1742-5689. doi: 10.1098/rsif.2013.0484. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3758008/.

[96] Michael Knudsen, Elisenda Feliu, and Carsten Wiuf. Exact Analysis of Intrinsic Qualitative Features of Phosphorelays using Mathematical Models. *arXiv:1109.5159*, September 2011. URL http://arxiv.org/abs/1109.5159.

[97] Elisenda Feliu and Carsten Wiuf. Enzyme-sharing as a cause of multistationarity in signalling systems. *Journal of The Royal Society Interface*, 9(71): 1224–1232, June 2012. ISSN 1742-5689, 1742-5662. doi: 10.1098/rsif.2011.0664. URL http://rsif.royalsocietypublishing.org/content/9/71/1224.

[98] Elisenda Feliu, Michael Knudsen, Lars N. Andersen, and Carsten Wiuf. An Algebraic Approach to Signaling Cascades with n Layers. *Bulletin of Mathematical Biology*, 74(1):45–72, January 2012. ISSN 0092-8240, 1522-9602. doi: 10.1007/s11538-011-9658-0. URL http://link.springer.com/article/10.1007/s11538-011-9658-0.

[99] J E Ferrell, Jr and E M Machleder. The biochemical basis of an all-or-none cell fate switch in Xenopus oocytes. *Science (New York, N.Y.)*, 280(5365):895–898, May 1998. ISSN 0036-8075.

[100] Qiang Zhang, Sudin Bhattacharya, and Melvin E. Andersen. Ultrasensitive response motifs: basic amplifiers in molecular signalling networks. *Open Biology*, 3(4), April 2013. ISSN , 2046-2441. doi: 10.1098/rsob.130031. URL http://rsob.royalsocietypublishing.org/content/3/4/130031.

[101] P Cotter. Phosphorelay control of virulence gene expression in Bordetella. *Trends in Microbiology*, 11(8):367–373, August 2003. ISSN 0966842X. doi: 10.1016/S0966-842X(03)00156-2. URL http://linkinghub.elsevier.com/retrieve/pii/S0966842X03001562.

[102] D Burbulys, K a Trach, and J a Hoch. Initiation of sporulation in B. subtilis is controlled by a multicomponent phosphorelay. *Cell*, 64(3):545–52, February 1991. ISSN 0092-8674. URL http://www.ncbi.nlm.nih.gov/pubmed/1846779.

[103] F Posas, S M Wurgler-Murphy, T Maeda, E a Witten, T C Thai, and H Saito. Yeast HOG1 MAP kinase cascade is regulated by a multistep phosphorelay mechanism in the SLN1-YPD1-SSK1 "two-component" osmosensor. *Cell*, 86 (6):865–75, September 1996. ISSN 0092-8674. URL http://www.ncbi.nlm.nih.gov/pubmed/8808622.

[104] Tatsuo Shibata and Koichi Fujimoto. Noisy signal amplification in ultrasensitive signal transduction. *Proceedings of the National Academy of Sciences of the United States of America*, 102(2):331–6, January 2005. ISSN 0027-8424. doi: 10.1073/pnas.0403350102. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=544281&tool=pmcentrez&rendertype=abstract.

[105] Takeshi Yoshida, Sheng jian Cai, and Masayori Inouye. Interaction of EnvZ, a sensory histidine kinase, with phosphorylated OmpR, the cognate response regulator. *Molecular microbiology*, 46(5):1283–1294, December 2002. ISSN 0950-382X.

[106] Gaurav Dogra, Frauke G. Purschke, Verena Wagner, Martin Haslbeck, Thomas Kriehuber, Jonathan G. Hughes, Maxwell L. Van Tassell, Crystal Gilbert, Melanie Niemeyer, W. Keith Ray, Richard F. Helm, and Birgit E. Scharf. Sinorhizobium meliloti CheA Complexed with CheS Exhibits Enhanced Binding to CheY1, Resulting in Accelerated CheY1 Dephosphorylation. *Journal of Bacteriology*, 194(5):1075–1087, March 2012. ISSN 0021-9193, 1098-5530. doi: 10.1128/JB.06505-11. URL http://jb.asm.org/content/194/5/1075.

[107] Dimitris Georgellis, Ohsuk Kwon, Peter De Wulf, and E. C. C. Lin. Signal Decay through a Reverse Phosphorelay in the Arc Two-component Signal Transduction System. *Journal of Biological Chemistry*, 273(49):32864–32869, December 1998. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.273.49.32864. URL http://www.jbc.org/content/273/49/32864.

[108] Sheng Jian Cai and Masayori Inouye. EnvZ-OmpR Interaction and Osmoregulation in Escherichia Coli. *Journal of Biological Chemistry*, 277(27):24155–24161, July 2002. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M110715200. URL http://www.jbc.org/content/277/27/24155.

[109] James E. Ferrell Jr. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends in Biochemical Sciences*, 21(12):460–466, December 1996. ISSN 0968-0004. doi: 10.1016/S0968-0004(96)20026-X. URL http://www.sciencedirect.com/science/article/pii/S096800049620026X.

[110] Joseph R Pomerening, Eduardo D Sontag, and James E Ferrell, Jr. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nature cell biology*, 5(4):346–351, April 2003. ISSN 1465-7392. doi: 10.1038/ncb954.

[111] Attila Becskei, Bertrand Séraphin, and Luis Serrano. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response

conversion. *The EMBO Journal*, 20(10):2528–2535, May 2001. ISSN 0261-4189, 1460-2075. doi: 10.1093/emboj/20.10.2528. URL http://emboj.embopress.org/content/20/10/2528.

[112] Ertugrul M. Ozbudak, Mukund Thattai, Han N. Lim, Boris I. Shraiman, and Alexander van Oudenaarden. Multistability in the lactose utilization network of Escherichia coli. *Nature*, 427(6976):737–740, February 2004. ISSN 0028-0836. doi: 10.1038/nature02298. URL http://www.nature.com/nature/journal/v427/n6976/abs/nature02298.html.

[113] Nick I. Markevich, Jan B. Hoek, and Boris N. Kholodenko. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades. *The Journal of Cell Biology*, 164(3):353–359, February 2004. ISSN 0021-9525, 1540-8140. doi: 10.1083/jcb.200308060. URL http://jcb.rupress.org/content/164/3/353.

[114] Fernando Ortega, José L Garcés, Francesc Mas, Boris N Kholodenko, and Marta Cascante. Bistability from double phosphorylation in signal transduction. Kinetic and structural requirements. *The FEBS journal*, 273(17):3915–3926, September 2006. ISSN 1742-464X. doi: 10.1111/j.1742-4658.2006.05394.x.

[115] T S Gardner, C R Cantor, and J J Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, January 2000. ISSN 0028-0836. doi: 10.1038/35002131.

[116] Călin C. Guet, Michael B. Elowitz, Weihong Hsing, and Stanislas Leibler. Combinatorial Synthesis of Genetic Networks. *Science*, 296(5572):1466–1470, May 2002. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1067407. URL http://www.sciencemag.org/content/296/5572/1466.

[117] Ellen C O'Shaughnessy, Santhosh Palani, James J Collins, and Casim A Sarkar. Tunable signal processing in synthetic MAP kinase cascades. *Cell*, 144(1):119–131, January 2011. ISSN 1097-4172. doi: 10.1016/j.cell.2010.12.014.

[118] Liming Wang and Eduardo D. Sontag. On the number of steady states in a multiple futile cycle. *Journal of Mathematical Biology*, 57(1):29–52, November 2007. ISSN 0303-6812, 1432-1416. doi: 10.1007/s00285-007-0145-z. URL http://link.springer.com/article/10.1007/s00285-007-0145-z.

[119] Matthew Thomson and Jeremy Gunawardena. Unlimited multistability in multisite phosphorylation systems. *Nature*, 460(7252):274–277, July 2009. ISSN 0028-0836. doi: 10.1038/nature08102. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2859978/.

[120] Katharina Holstein, Dietrich Flockerzi, and Carsten Conradi. Multistationarity in Sequential Distributed Multisite Phosphorylation Networks. *Bulletin of Mathematical Biology*, 75(11):2028–2058, September 2013. ISSN 0092-8240, 1522-9602. doi: 10.1007/s11538-013-9878-6. URL http://link.springer.com/article/10.1007/s11538-013-9878-6.

[121] Munia Amin, Steven L. Porter, and Orkun S. Soyer. Split Histidine Kinases Enable Ultrasensitivity and Bistability in Two-Component Signaling Networks. *PLoS Comput Biol*, 9(3):e1002949, March 2013. doi: 10.1371/journal.pcbi.1002949. URL http://dx.doi.org/10.1371/journal.pcbi.1002949.

[122] Rebecca B. Hoyle, Daniele Avitabile, and Andrzej M. Kierzek. Equation-Free Analysis of Two-Component System Signalling Model Reveals the Emergence of Co-Existing Phenotypes in the Absence of Multistationarity. *PLoS Comput Biol*, 8(6):e1002396, June 2012. doi: 10.1371/journal.pcbi.1002396. URL http://dx.doi.org/10.1371/journal.pcbi.1002396.

[123] David Dubnau and Richard Losick. Bistability in bacteria. *Molecular microbiology*, 61(3):564–572, August 2006. ISSN 0950-382X. doi: 10.1111/j.1365-2958.2006.05249.x.

[124] Hédia Maamar, Arjun Raj, and David Dubnau. Noise in Gene Expression Determines Cell Fate in Bacillus subtilis. *Science*, 317(5837):526–529, July 2007. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1140818. URL http://www.sciencemag.org/content/317/5837/526.

[125] Michael Y. Galperin. A census of membrane-bound and intracellular signal transduction proteins in bacteria: Bacterial IQ, extroverts and introverts. *BMC Microbiology*, 5(1):1–19, December 2005. ISSN 1471-2180. doi: 10.1186/1471-2180-5-35. URL http://link.springer.com/article/10.1186/1471-2180-5-35.

[126] Weiwen Zhang and Liang Shi. Distribution and evolution of multiple-step phosphorelay in prokaryotes: lateral domain recruitment involved in the formation of hybrid-type histidine kinases. *Microbiology*, 151(7):2159–2173, July 2005. ISSN 1350-0872, 1465-2080. doi: 10.1099/mic.0.27987-0. URL http://mic.sgmjournals.org/content/151/7/2159.

[127] Gheorghe Craciun and Martin Feinberg. Multiple equilibria in complex chemical reaction networks: I. the injectivity property. *SIAM Journal on Applied Mathematics*, 65:1526–1546, 2005.

[128] Murad Banaji and Casian Pantea. Some results on injectivity and multistationarity in chemical reaction networks. arXiv e-print 1309.6771, Middlesex University London, September 2013. URL http://arxiv.org/abs/1309.6771.

[129] Murad Banaji and G. Craciun. Graph-theoretic approaches to injectivity and multiple equilibria in systems of interacting elements. *Communications in Mathematical Sciences*, 7(2):867–900, December 2009. ISSN 1539-6746. URL http://eprints.port.ac.uk/2922/.

[130] Badal Joshi and Anne Shiu. Atoms of multistationarity in chemical reaction networks. *arXiv:1108.5238 [math, q-bio]*, August 2011. URL http://arxiv.org/abs/1108.5238.

[131] David C. Kurtz. A Sufficient Condition for All the Roots of a Polynomial To Be Real. *The American Mathematical Monthly*, 99(3):259–263, March 1992. ISSN 0002-9890. doi: 10.2307/2325063. URL http://www.jstor.org/stable/2325063.

[132] Steffen Waldherr. A guideline to model reduction by stoichiometric decomposition for biochemical network analysis. In *Proc. of the 21st International Symposium on Mathematical Theory of Networks and Systems pages:*, pages 490–495. Mathematical Theory of Networks and Systems, Non-KU Leuven Association publications, July 2014.

[133] C. Wiuf and E. Feliu. Power-Law Kinetics and Determinant Criteria for the Preclusion of Multistationarity in Networks of Interacting Species. *SIAM Journal on Applied Dynamical Systems*, 12(4):1685–1721, January 2013. doi: 10.1137/120873388. URL http://epubs.siam.org/doi/abs/10.1137/120873388.

[134] Tao Long, Kimberly C Tu, Yufang Wang, Pankaj Mehta, N. P Ong, Bonnie L Bassler, and Ned S Wingreen. Quantifying the Integration of Quorum-Sensing Signals with Single-Cell Resolution. *PLoS Biol*, 7(3):e1000068, March 2009. doi: 10.1371/journal.pbio.1000068. URL http://dx.doi.org/10.1371/journal.pbio.1000068.

[135] Kaneyoshi Yamamoto, Kiyo Hirao, Taku Oshima, Hirofumi Aiba, Ryutaro Utsumi, and Akira Ishihama. Functional Characterization in Vitro of All Two-component Signal Transduction Systems from Escherichia coli. *Journal of Biological Chemistry*, 280(2):1448–1456, January 2005. ISSN 0021-9258, 1083-351X. doi: 10.1074/jbc.M410104200. URL http://www.jbc.org/content/280/2/1448.

[136] Kirsten R Guckes, Maria Kostakioti, Erin J Breland, Alice P Gu, Carrie L Shaffer, Charles R Martinez, 3rd, Scott J Hultgren, and Maria Hadjifrangiskou. Strong cross-system interactions drive the activation of the QseB response regulator in the absence of its cognate sensor. *Proceedings of the National Academy of Sciences of the United States of America*, 110(41):16592–16597, October 2013. ISSN 1091-6490. doi: 10.1073/pnas.1315320110.

[137] Irene L. G. Newton and Seth R. Bordenstein. Correlations Between Bacterial Ecology and Mobile DNA. *Current microbiology*, 62(1):198–208, January 2011. ISSN 0343-8651. doi: 10.1007/s00284-010-9693-3. URL `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3006647/`.

[138] Varun B. Kothamachu, Elisenda Feliu, Luca Cardelli, and Orkun S. Soyer. Unlimited multistability and boolean logic in microbial signalling. *Journal of The Royal Society Interface*, 12(108), June 2015. ISSN 1742-5689. doi: 10.1098/rsif.2015.0234.

[139] Alex Mira, Howard Ochman, and Nancy A. Moran. Deletional bias and the evolution of bacterial genomes. *Trends in Genetics*, 17(10):589–596, 2016/06/07 2001. doi: 10.1016/S0168-9525(01)02447-7. URL `http://dx.doi.org/10.1016/S0168-9525(01)02447-7`.

[140] Mohamed Barakat, Philippe Ortet, and David E Whitworth. P2cs: a database of prokaryotic two-component systems. *Nucleic Acids Res*, 39(Database issue): D771–6, Jan 2011. ISSN 1362-4962 (Electronic); 0305-1048 (Linking). doi: 10.1093/nar/gkq1023.

[141] Mohamed Barakat, Philippe Ortet, Cécile Jourlin-Castelli, Mireille Ansaldi, Vincent Méjean, and David E. Whitworth. P2cs: a two-component system resource for prokaryotic signal transduction research. *BMC Genomics*, 10 (1):315, July 2009. ISSN 1471-2164. doi: 10.1186/1471-2164-10-315. URL `http://www.biomedcentral.com/1471-2164/10/315/abstract`.

[142] Elhanan Borenstein, Martin Kupiec, Marcus W. Feldman, and Eytan Ruppin. Large-scale reconstruction and phylogenetic analysis of metabolic environments. *Proceedings of the National Academy of Sciences*, 105(38):14482–14487, September 2008. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.0806162105. URL `http://www.pnas.org/content/105/38/14482`.

[143] Kim D Pruitt, Tatiana Tatusova, and Donna R Maglott. Ncbi reference sequence (refseq): a curated non-redundant sequence database of genomes,

transcripts and proteins. *Nucleic Acids Research*, 33(Database Issue):D501–D504, 01 2005. doi: 10.1093/nar/gki025. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC539979/.

[144] Kristin Wuichet, Brian J Cantwell, and Igor B Zhulin. Evolution and phyletic distribution of two-component signal transduction systems. *Current opinion in microbiology*, 13(2):219–225, April 2010. ISSN 1879-0364. doi: 10.1016/j.mib.2009.12.011.

[145] Peter J. A. Cock and David E. Whitworth. Evolution of Prokaryotic Two-Component System Signaling Pathways: Gene Fusions and Fissions. *Molecular Biology and Evolution*, 24(11):2355–2357, November 2007. ISSN 0737-4038, 1537-1719. doi: 10.1093/molbev/msm170. URL http://mbe.oxfordjournals.org/content/24/11/2355.

[146] David E Whitworth and Peter J A Cock. Evolution of prokaryotic two-component systems: insights from comparative genomics. *Amino acids*, 37(3):459–466, September 2009. ISSN 1438-2199. doi: 10.1007/s00726-009-0259-2.

[147] Kristin K. Koretke, Andrei N. Lupas, Patrick V. Warren, Martin Rosenberg, and James R. Brown. Evolution of Two-Component Signal Transduction. *Molecular Biology and Evolution*, 17(12):1956–1970, December 2000. ISSN 0737-4038, 1537-1719. URL http://mbe.oxfordjournals.org/content/17/12/1956.

[148] Emily J. Capra and Michael T. Laub. Evolution of Two-Component Signal Transduction Systems. *Annual Review of Microbiology*, 66(1):325–347, 2012. doi: 10.1146/annurev-micro-092611-150039. URL http://dx.doi.org/10.1146/annurev-micro-092611-150039.

[149] Eric Alm, Katherine Huang, and Adam Arkin. The Evolution of Two-Component Systems in Bacteria Reveals Different Strategies for Niche Adaptation. *PLoS Comput Biol*, 2(11):e143, November 2006. doi: 10.1371/journal.pcbi.0020143. URL http://dx.plos.org/10.1371/journal.pcbi.0020143.

[150] Ana Conesa and Stefan Gotz. Blast2go: A comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics*, page 619832, 2008. ISSN 1687-5370 (Print); 1687-5389 (Linking). doi: 10.1155/2008/619832.

[151] Ana Conesa, Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Manuel Talón, and Montserrat Robles. Blast2go: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics (Oxford, England)*, 21(18):3674–3676, September 2005. ISSN 1367-4803. doi: 10.1093/bioinformatics/bti610.

[152] Stefan Götz, Juan Miguel García-Gómez, Javier Terol, Tim D. Williams, Shivashankar H. Nagaraj, María José Nueda, Montserrat Robles, Manuel Talón, Joaquín Dopazo, and Ana Conesa. High-throughput functional annotation and data mining with the Blast2go suite. *Nucleic Acids Research*, 36(10): 3420–3435, June 2008. ISSN 0305-1048, 1362-4962. doi: 10.1093/nar/gkn176. URL `http://nar.oxfordjournals.org/content/36/10/3420`.

[153] H Fujii, T Shimada, T Eendo, T Shimizu, and M Omura. 29,228 Citrus ESTs-Collection And Analysis Toward The Functional Genomics Phase. *Plant & Animal Genomes XIV Conference*, 2006.

[154] S F Altschul, W Gish, W Miller, E W Myers, and D J Lipman. Basic local alignment search tool. *J Mol Biol*, 215(3):403–410, Oct 1990. ISSN 0022-2836 (Print); 0022-2836 (Linking). doi: 10.1016/S0022-2836(05)80360-2.

[155] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics*, 25 (1):25–29, May 2000. ISSN 1061-4036. doi: 10.1038/75556.

[156] The Gene Ontology Consortium. Gene ontology consortium: going forward. *Nucleic Acids Research*, 43(D1):D1049–D1056, 01 2015. URL `http://nar.oxfordjournals.org/content/43/D1/D1049.abstract`.

[157] Nils Blüthgen, Karsten Brand, Branka Cajavec, Maciej Swat, Hanspeter Herzel, and Dieter Beule. Biological profiling of gene groups utilizing Gene Ontology. *Genome Informatics. International Conference on Genome Informatics*, 16(1): 106–115, 2005. ISSN 0919-9454.

[158] Beate Krueger, Torben Friedrich, Frank Förster, Jörg Bernhardt, Roy Gross, and Thomas Dandekar. Different evolutionary modifications as a guide to rewire two-component systems. *Bioinformatics and biology insights*, 6:97–128, 2012. ISSN 1177-9322. doi: 10.4137/BBI.S9356.

[159] Jeffrey M Skerker, Barrett S Perchuk, Albert Siryaporn, Emma A Lubin, Orr Ashenberg, Mark Goulian, and Michael T Laub. Rewiring the Specificity of Two-Component Signal Transduction Systems. *Cell*, pages 1043–1054, 2008. doi: 10.1016/j.cell.2008.04.040.

[160] Weston R Whitaker, Stephanie A Davis, Adam P Arkin, and John E Dueber. Engineering robust control of two-component system phosphotransfer using

modular scaffolds. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44):18090–18095, October 2012. ISSN 1091-6490. doi: 10.1073/pnas.1209230109.

[161] Anselm Levskaya, Aaron A. Chevalier, Jeffrey J. Tabor, Zachary Booth Simpson, Laura A. Lavery, Matthew Levy, Eric A. Davidson, Alexander Scouras, Andrew D. Ellington, Edward M. Marcotte, and Christopher A. Voigt. Synthetic biology: Engineering Escherichia coli to see light. *Nature*, 438(7067):441–442, November 2005. ISSN 0028-0836. doi: 10.1038/nature04405. URL `http://www.nature.com/nature/journal/v438/n7067/full/nature04405.html`.